

Towards a mild nuclear winter?

Two years of introspection about the climatic consequences of a nuclear war come to a climax in Berne next week. The outlook seems less chilly than was thought.

NEXT week's general assembly of the International Council of Scientific Unions (ICSU) will have a muddy issue on its hands, that of the reality or otherwise of what has become known in the past few years as the nuclear winter. The occasion is the formal delivery of the report on the consequences of a nuclear war commissioned by ICSU's off-shoot SCOPE (Scientific Committee on Problems of the Environment) from an *ad hoc* working group called ENUWAR. The formal report of this two-year study has been published in bits and pieces since the beginning of this year. Those going to Berne (where the meeting is to be held), or fearful of being mystified by reports of the occasion, should be sure to read the inconclusive but illuminating exchange of opinions on the subject which appears in the Fall issue of the journal *Foreign Affairs* (page 163 *et seq.*). Meanwhile, a summary of where things now stand may be of some assistance.

The first formal statement of the case for believing that a substantial nuclear exchange between the two major nuclear powers would have serious climatic consequences is due to R.P. Turco *et al.* (*Science* **222**, 1283–1292 1982) and was based on calculations made with a one-dimensional model of the radiative balance in an atmosphere laden with smoke from all the fires set by the explosion of a vast arsenal of nuclear weapons. Provided the amount of smoke was large enough, temperatures in northern mid-latitudes were expected to fall by an average of 15 degrees Celsius for some months. Amid expressions of scepticism about the accuracy of the result, a group from NCAR (National Center for Atmospheric Research) at Boulder, Colorado (Covey *et al.* *Nature* **308**, 21 1985), described the results of a three-dimensional study leading to a similar but less extreme result (which was only to be expected). The SCOPE report, whose summary was made public at the beginning of the year, concluded that nothing that had happened during the intervening interval had "lessened the probability that a major nuclear exchange would cause severe environmental effects".

Cooling

There is now no substantial dissent from the opinion that fire-raising on the vast scale made possible by an exchange of nuclear weapons would load the atmosphere with so much smoke that there would be climatic effects of some kind. The question on which the argument now turns is the degree to which it is reasonable to suppose that the surface of the Earth would be cooled, and for how long. The SCOPE report might have helped to settle the issue more decisively had it had been published complete, and if it had been written by a single author, preferably one not identified with a declared position.

The argument in *Foreign Affairs* has taken a different tack, that of the strategic and political implications of the threat of a nuclear winter. Dr Carl Sagan set the ball rolling (in the Winter, 1983/84 issue) with the argument that the threat of nuclear winter implies, among other things, that the nuclear powers should aim at negotiating a reduction of their nuclear arsenals to the point at which they each have less than one per cent of their present stocks; this would be enough for "minimal deterrence", but not enough to cause a nuclear winter. In the summer of this year, and on the basis of a string of calculations at NCAR in the

interval, Covey *et al.* begged leave to dissent, saying that the temperature might be reduced by only a third of the amount first calculated, although everybody agrees that there would be some deterioration of the climate in the hemisphere in which a nuclear exchange took place. (The comparison of the results of calculations with the two different kinds of models is not straightforward, so that little should be made of the discrepancy reported.) The interest of this development is that Covey *et al.* argue for a nuclear "autumn" rather than a winter, a climatic perturbation not severe enough to cause widespread cataclysm outside the regions in which the bombs fall. Dr Carl Sagan, one of the authors of the original account, dissents, insisting that not much has changed.

Trigger-happy

Given the persisting uncertainties, it is understandable that some should ask impatiently what all the fuss can be about. On the principle that it is no better to be hung for a sheep than a lamb, it might be argued that contemplation of the consequences of a nuclear war would more prudently be based on nuclear winter, not merely nuclear autumn. But that is a short-sighted argument. Covey *et al.* in their rejoinder, argue cogently that a 99 per cent reduction of the superpowers' nuclear arsenals would lead to an exceedingly dangerous state of affairs if it were not accompanied by other safeguards, hitherto unspecified. Nuclear powers would be compelled to aim their surviving warheads at cities, not military targets, and would be made trigger-happy by the constant fear that the other fellow might strike first. (Sagan complains that Mr Richard Perle, US Assistant Secretary of Defense, has used the threat of nuclear winter to justify the Strategic Defense Initiative, but Perle is logically correct.)

George W. Rathjens and Ronald H. Siegal from the Massachusetts Institute of Technology take the argument further, suggesting that even Covey *et al.* may have overestimated the climatic consequences of smoke in the atmosphere and making the cogent point that preoccupation with the early extreme predictions can only distract people of goodwill from more urgent tasks in international relations, reducing nuclear arsenals by a half, for example.

What, in these circumstances, can ICSU say about the ENUWAR report? It might start by acknowledging that the authors of the original hypothesis of nuclear winter (which include Crutzen and Birks as well as Turco *et al.*) have hit on an interesting problem which is likely to run and run. As it happens, ICSU has an ambition to launch a programme of research on the global environment, intended to provide a kind of baseline for the later assessment of man-made effects, among which nuclear wars would certainly qualify. In due course, such a programme might provide a better basis for the kind of assessment that the nuclear climatologists have been attempting, although the persisting uncertainties are chiefly those inherent in climatic models at their present stage of development and the inevitable uncertainty about the precise pattern in which warheads would be scattered in a real nuclear war. Meanwhile, the letter from Mr Russell Seitz on page 116 of this issue (oddly enough, echoed in some



places word for word by a paper to be delivered next week in Berne by Academician G.S. Golitsyn of the Institute of Atmospheric Physics in Moscow) shows that there may be a rich vein of historical research yet to be tapped. ICSU itself might think of issuing a temporizing opinion on the matter, one that draws attention to the uncertainties running through all the arguments so far advanced to calculate the consequences of a nuclear war, to work that remains undone and, recognizing that the stated purpose of deterrence is to avoid nuclear war, urges those with good intentions to direct them at practical tasks. □

Academics in clover?

Harvard prompts the question whether universities are being diverted from academic ways.

HARVARD University is rightly proud of its place in modern scholarship as well as of its role in formulating and defending the concept of the autonomous university. So the element of hoopla in last week's celebration of Harvard's 350th anniversary (see page 103) is forgivable, and was probably unavoidable. How else, for that matter, does a great institution remind itself of past achievements, and gather the resolve to make the future as bright as possible? Yet there are obvious dangers, of which outstanding institutions are well aware. Especially because the reputation of a university rests on its corporate intellectual quality, there is always a narrow gap between others' admiration and resentment. In most places, the most successful universities are regarded as bastions of privilege by academics elsewhere, just as universities as a whole may be tarred with the same brush in the more general opinion. In the United States there are now, to the common good fortune, a dozen institutions other than Harvard with comparable reputations, mostly won by quite different means.

Yet Harvard has a problem, one that must peculiarly afflict an institution whose cry has always been the primacy of scholarship and its accompaniments, teaching and research. The difficulty was well described last week by Dr Derek Bok, the university's president: inevitably, as public funds for the support of research shrink, or are spread over a growing throng of claimants, academics must turn to other sources both for funds and, ultimately, for approval. Industrial partnerships of various kinds are now the fashion. Bok might have added that, as in other ways, academics at the more successful institutions are the better able to lay claim to these alternative sources of support.

The Harvard faculty restrains its members from carrying out militarily classified research, but the restraints on commercial work are less explicit. The difficulty becomes a dilemma because, as Bok acknowledged last week, there is a case for asking that outstanding institutions and their academics can usefully contribute to wider social objectives by assisting industrial innovation. The question is where to draw the line between outside commitments which are productive and even stimulating for those concerned and commitments which are distracting and a threat to scholarship. Harvard is obviously worried that the line may already be blurred.

For what it is worth, the problem is not unique to Harvard nor intractable. The first need is that its nature, and the danger that it poses, should be recognized for what they are. The most serious danger is that too close an involvement with commercial activities is intellectually and emotionally distracting, and a thief of people's time. The extreme case is that when academics have some kind of management responsibility for extramural enterprises; then, if things go wrong, their fiducial responsibilities may require them to abandon their responsibilities to students and to research. Other dangers are more sordid, to do with the money that academics can now earn by energetic moonlighting. A particularly insidious danger is that academics might come to accept as tests of achievement standards which have nothing to do with scholarship, thereby compromising the distinctiveness

of the university as an institution and the reasons for its continuing existence.

The safeguards are also obvious. What academics do extramurally should be made clear to their colleagues, whose opinions of what is permissible should carry at least as much weight as that of would-be entrepreneurs. Attempts to control the time that people spend on outside activities are inevitable but less effective than they appear; who can control what goes on inside a person's head? A better general solution would be to require that extramural workers should share their rewards with their universities, at the institutional or departmental level.

In the past few years, the Harvard faculty has dealt intelligently with a number of cases of this kind thrown up by the emergence of new techniques in biotechnology. But now the field of opportunity is widening. In the circumstances, the most urgent need is for a more explicit way of regulating these affairs. That Harvard should have been among the first publicly to acknowledge the danger is not a sign that it is far along the slippery slope, but that it still knows what universities are for. Research, of course. Teaching, more earnestly than at present, even at Harvard. But also the constructive irreverence made possible by the conjunction of courageous people and independence. Universities everywhere can boast of the first; Harvard's good fortune is that it still enjoys the second. □

Hijacking riposte

Another hijacked aircraft at Karachi last week suggests the need for more deliberate safeguards.

HIJACKING aircraft is not merely criminal terrorism but an increasingly serious limitation of people's freedom to move about the surface of the Earth. With each new advance in the technology of remote (and not so remote) sensing for explosives and firearms, airlines understandably require passengers to arrive even earlier at the check-in counters, increasing the proportion of each journey which constitutes dead time while adding substantially to the total cost. Yet the security people are the first to say that they are necessarily looking, without much success to judge from the lack of prosecutions for carrying weapons onto aircraft, for needles in a haystack. So might they not think of narrowing the search, in a field in which even innocent passengers must be judged suspect until they are cleared, by eliminating a substantial number of people from scrutiny?

Here is how it might be done. Part of the airport problem is to know the identity of travellers, so why not arrange that potential passengers may, if they wish, establish their identity in advance by telling some airline computer who they are and what they do? Whether the same person later turns up at the check-in counter could then be verified by one of the devices which the banks are contemplating for telling the identity of individuals; a person might, for example, record the ages of siblings, the nick-names of schoolteachers and a host of other trivial but idiosyncratic information, all of which might be checked by a few questions at random formulated by a computer. A few seconds at an airport desk might replace the tens of minutes now spent crawling through the security checks.

Only the airlines can tell how great the saving of time and effort would be, but it cannot but be substantial. The people most likely to volunteer their identities in such a way would most probably be the frequent travellers, but they are by definition a substantial part of the airline traffic. Irregular travellers would complain of being at a disadvantage, but a suitable system could be made to work for all who volunteer. But who says that it is a fair price to pay for the abatement of hijacking that there should be a huge data-bank filled with intimate information about people at random? Properly run, a system like this should not be an offence against civil liberties: of necessity, only information so trivial that imposters would never know it would be required. □

NASA schedule

Prospects still bad for space science launches

Washington

Dr Burton Edelson, associate administrator for space science and applications for the National Aeronautics and Space Administration (NASA), last week made public revised plans for space science missions in the aftermath of the Challenger accident. While the two-year delay to shuttle launches and the subsequent lower launch rate mean that many missions will be postponed or cancelled, Edelson is determined that science missions should be fairly represented when launches resume. He said it would be "ruinous" if NASA had to use money from its science budget to build the unmanned rockets now planned. Edelson also announced new initiatives to restore the enthusiasm of university researchers whose shuttle experiments have been delayed.

Edelson, who was addressing a meeting of the American Astronautical Society, said that NASA now plans some 63 shuttle missions between 1988 (when flights should resume) and 1992, although he added that some thought that schedule

analysis and sub-orbital research; for developing new instruments and small payloads; for using other NASA flight projects to give "piggyback" launches to science projects; and for developing new cooperative programmes with domestic and international partners.

Edelson said that eleven major Spacelab-type shuttle flights that were to have flown between 1987 and 1989 had been on the list for cancellation, but that four have now been identified as launch candidates (though only three are likely to go before

1990). These are the Earth Observation Mission, a joint effort with the European Space Agency, now renamed Atlas; Astro, an ultraviolet observatory; the International Microgravity Laboratory; and SLS-1, a dedicated Spacelab life sciences flight. But even those that do fly are likely to suffer delays of at least two and a half years. Between 8 and 10 materials science laboratory payloads and about 12 Spartan payloads will also now not fly before 1990.

Edelson also said that NASA's recent decision to use Expendable Launch Vehicles (unmanned rockets) as well as the shuttle means that as many as 8 of the 145 missions planned before 1992 could be launched this way. But he gave a stern warning that if NASA is to procure new unmanned rockets, it would need extra funds for the purpose. So far, Congress has ducked the issue. **Tim Beardsley**

Reproductive technology

Ethical guidelines announced

Washington

THE American Fertility Society this week published comprehensive guidelines on the uses of new reproductive technology, including *in vitro* fertilization and the freezing of embryos and gametes. The guidelines are the first attempt in the United States to achieve a broad professional consensus on the ethical issues raised by the new technology. The eleven-member ethics committee which drew them up includes many of the well-known names in the field, under the chairmanship of Howard Jones of Johns Hopkins University School of Medicine. Although the society cannot enforce compliance, its 10,000 members are likely to respect the guidelines.

The guidelines appear as a special supplement to the September 1986 issue of the society's journal *Fertility and Sterility*. The committee finds no overwhelming ethical objection to any of the reproductive techniques now being considered, but warns that some should still be considered as experimental and subject to special scrutiny by local institutional review boards. It asks the Fertility Society to press for an expanded national debate to include other groups, including the recently inaugurated Congressional biomedical ethics board.

The committee also supports research on the new techniques provided it is backed up with sound policy to protect individual freedom and to gain informed consent. For research on human pre-embryos (a term it advocates for the zygote up to the 8-cell stage), the committee advocates setting a time limit so that the pre-embryo is destroyed 14 days after fertilization. This is the limit voluntarily adopted in Britain and in Australia.

The committee makes a strong call for

the Department of Health and Human Services to abandon its present *de facto* moratorium on supporting research involving human *in vitro* fertilization or embryo transfer. The department's regulations forbid it from supporting any such research that has not been approved by its Ethics Advisory Board; but the board was disbanded in 1980, and its 1979 recommendation that *in vitro* fertilization research is acceptable under certain conditions has not been acted on. There are also statutory limitations imposed by congress prohibiting federal funds from being used for similar research, with the result that research is concentrated in private clinics and hospitals.

The committee rules that techniques such as *in vitro* fertilization and surrogate motherhood should be considered only when parents are unable to produce children and not merely for convenience. But although maintaining that a stable heterosexual couple provides the best environment for child-rearing, the committee would not exclude the use of the techniques for others who may want children.

Besides simple *in vitro* fertilization, the committee considers that the use of donor pre-embryos and the cryopreservation of sperm would be acceptable. But it says that uterine lavage as means of obtaining pre-embryos for transfer should be regarded as a clinical experiment and so subject to full research approval requirements; similarly, cryopreservation of pre-embryos and eggs is seen as premature for general use but a valid topic for research. Surrogate motherhood, about which the committee has "serious reservations", should also be considered experimental. But the committee stresses that its recommendations should be subject to change in the light of public debate. **Tim Beardsley**



Priority but still a distant prospect. The Hubble telescope.

optimistic. This represents a major reduction, since before the Challenger accident, 145 missions were planned by 1992. Of the 80 lost flights, 31 were to have been primarily space science missions, now only 19 science missions will fly before 1992.

Acknowledging that his news was not good, Edelson restated NASA's commitment to ensure that space science missions should occupy at least a third of all shuttle flights. He also said that launching the Hubble Space Telescope remains a top priority.

Edelson's programme ("some tens of millions of dollars") for maintaining the health of the university research community includes more support for data

Australian budget

Science bucking the trend?

Sydney

THIS year's Australian budget is being hailed as a victory for science and for the Minister for Science, Mr Barry Jones. Adapting to harsh new economic realities, across-the-board cuts were the order of the day, including A\$500 million pared from welfare, and the reintroduction of limited fees for tertiary education. But science slipped past the razor almost unscathed.

The grim state of Australia's economy is reflected in the plunge in the value of the dollar, and a massive trade deficit due to low prices for Australia's traditional exports of farm produce and minerals and a dismal performance by manufacturing industry. The Treasurer, Mr Paul Keating, warned that if trade problems were not solved soon the country would be forced to take "banana republic" style measures.

Despite the smooth ride for science, the Commonwealth Science and Industrial Research Organization (CSIRO), taking two-thirds of the budget, must absorb about a 1.5 per cent cut in real terms in funds for salaries and operating costs. This is due to inflation running at 8 per cent and A\$5 million earmarked for a retirement scheme to allow young researchers to move into areas thought to be of special significance for Australia's future. Six areas have been chosen: computers, information technology, manufacturing technology, raw materials processing, space technology, and (human) nutrition.

The best news is for the Australian Research Grants Scheme (ARGS) which was blessed with a 10 per cent increase in funds in real terms. The government has a commitment to strengthen the role of ARGs which could eventually see it take on a role like that of the National Science Foundation in the United States.

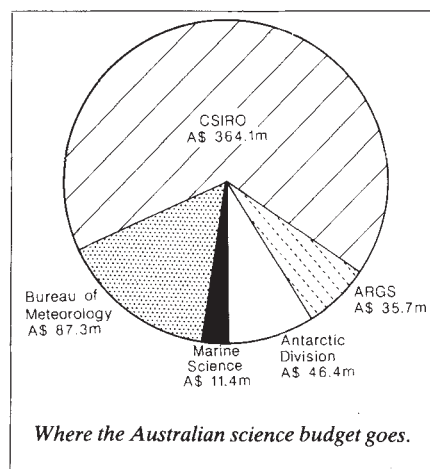
Funds for the Bureau of Meteorology are just keeping pace with inflation. But the bureau does have A\$5.6 million this year to pay for equipment upgrading and construction costs, compared with A\$3.6 million last year. High priority is being given to equipment to warn of severe weather such as tropical cyclones and the floods which struck Sydney in August.

The Antarctic Division's funds are up by A\$4.3 million but ship charter costs will rise by A\$4.5 million in 1986/87 because of the fall in the value of the dollar. An additional bonus of A\$4.5 million this year, following on from A\$5.0 million last year, will help continue the building of new bases, which will last into the next century.

Marine science turns out a loser. Professor Jörg Inberger, President of the Australian Marine Science Association described the budget as a catastrophe.

Grants under the Marine Science and Technology Scheme have been cut from A\$4.0 million in 1985-86 to A\$3.8 million. He is particularly disappointed when increases of 25 to 30 per cent are needed just to keep existing programmes going due to inflation, the falling dollar, and the need for increasingly sophisticated and expensive research tools. In part the funds saved will go to fulfil promises that ARGs funds would be increased. The Australian Institute of Marine Science received just sufficient to keep up with inflation.

Professor Inberger also finds no joy in his role as chairman of the ARGs subcommittee for Engineering and Applied Science. His committee submitted a request for A\$12 million to fund research that they believed essential. They received A\$5.5 million. Moreover the government has earmarked A\$1.2 million of the A\$5.5



Where the Australian science budget goes.

million for large items of equipment because it believes that funding fewer, but larger, projects will make more productive use of the money. But this strategy means there will be no increase in the number of projects funded or scientists supported.

Charles Morgan

Science policy

US lobby group still struggling

Washington

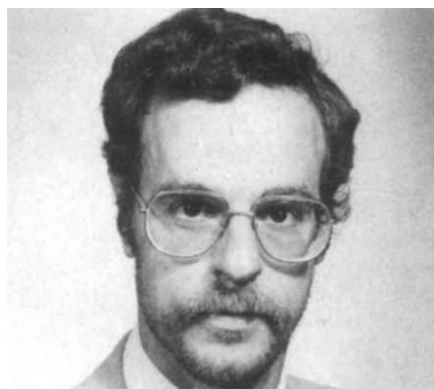
THE National Coalition for Science and Technology (CST), the only unabashed US lobbying group claiming to represent science on Capitol Hill, is facing an uncertain future. The five-year-old organization has already once been rescued from the brink of bankruptcy, but unless another source of revenue is found soon, it may have to close its doors for good.

The coalition's executive director, Philip Speser, is forthright about its aims: to make sure that the interests of the scientific community are heard loud and clear in the corridors of power. The work, he says, consists largely of "knocking on doors". Because CST is a registered lobbying group and not a charity (unlike most other scientific organizations that attempt discreetly to pull strings in Washington), CST can openly express its views on key members of Congress.

The response of the scientific community has, however, been less than enthusiastic. Speser thinks this stems from scientists' deeply-ingrained distaste for direct involvement in the political process. Scientists prefer to make their case through "non-lobbying lobbyists" — the professional societies that wait politely to be asked their opinions at congressional hearings.

Individual membership of the coalition (at a modest \$35 per year) is down to less than 300, from a peak of around 500 a year ago. About 30 corporations and institutions have signed up, but they are not enough to secure the coalition's future.

The coalition is run by Foresight Science and Technology Inc., a government



CST director Philip Speser: still knocking on doors.

relations and consulting company of which Speser is president; it also handles several other non-profit organizations, making its profits from, among other things, books on how to secure government contracts for small businesses.

Recent activities include the publication regular newsletters, and congressional breakfasts where scientists can meet political decision-makers. The coalition has tried to maximize its effect by concentrating on specific issues that have not been taken up by more influential groups. But it is far from well-known, and staff on several science committees in Congress say they have never heard of it.

Speser says he has the behind-the-scenes support of senior science mandarins in the government who are encouraging him to keep the coalition afloat, and that the coalition was a significant, if not revolutionary, influence on the 1987 science budget.

Tim Beardsley

Biotechnology

Doubts over new degree courses

Washington

WHILE biotechnology has long since won the confidence of investors, it is only just beginning to appear as a degree subject in its own right. This fall, several two- and four-year programmes will emerge. Lacking precedent or prototype, their curricula are intended to meet industry's needs. But industry, which is not at present grappling with a labour shortage, is greeting some of the programmes with ambivalence.

Although programme designs vary some common threads run through them all. They tend to be long on chemistry, mathematics and biology, but short on electives. They will be counted among the most difficult degrees offered at most institutions. They will also be expensive: one state-subsidized programme costs about \$10,000 per student for one semester, but charges only \$780 tuition.

Above all, biotechnology programmes emphasize technical proficiency and hands-on experience. Students will become adept with protocols and equipment they would not ordinarily encounter until graduate school. "To my mind, biotechnology isn't a discipline in itself", says Jean Brenchley, who heads the new biotechnology institute at Pennsylvania State University. "I see it building upon traditional disciplines, and the techniques are what serve as the focal point."

The university recruited Brenchley from Genex Corporation in 1984 to direct the institute's research activities, but students encouraged her to devise a degree programme as well. Penn State's programme brings many different departments under the umbrella of the \$7 million biotechnology facility which will be complete this winter. Engineers will be expected to learn about the metabolic requirements of microbes, and molecular biologists will be taught to think about large-scale production.

Brenchley's curriculum represents one of the two approaches to biotechnology training. Her graduates will be groomed for PhD programmes or research and development laboratories with starting salaries around \$25,000. In contrast, other academics are aiming at a still-nascent niche for the production and quality control technician who would require less theoretical education and would ultimately receive a smaller salary. In order to assess the opportunity for such positions, 10 two-year colleges in the mid-west last year commissioned a report from the Center for Occupational Research and Development in Waco, Texas.

The projections were rather optimistic. Pooling the responses of 86 companies in the industry, the report predicts that the number of US biotechnicians will increase

by 148 per cent by 1995. That means 1,000 new job openings per year for people with associate degrees in biotechnology.

Industry, while expressing enthusiasm for bachelors' programmes, seems sceptical that two years in college can provide the knowledge necessary to manage complex instrumentation and sensitive organisms. Michael Montague, director of Monsanto's biological sciences division calls the Waco centre's 1,000 a year projection "definitely a guess". Others fear that technicians who lack theoretical understanding may also lack the flexibility to adjust to a rapidly-evolving technology. And they acknowledge that two-year programmes bear a stigma in an industry where many of the executives are PhDs.

Associate degree programmes may face a barrier more intractable than industry prejudice — cost. The equipment, supplies and faculty necessary to teach state-of-the-art techniques can carry prohibitive price tags. Some more seasoned four-year programmes have hit upon money saving solutions.

Industry contributions, for example, can take the edge off equipment expenditure. At the Minor Institute in Chazy, New York, many of the laboratory expendables for an intensive one-semester programme in biotechnology are gifts from the manufacturers, who hope to "hook" students on their brands. Academic coordinator William Graziadei, who forged the link between Minor and Plattsburgh State, developed a network of adjunct professors that brings 35 industry and academic experts to the remote university town each year, yet keeps the programme's payroll down to a minimum.

In Seattle, Washington, local colleges have established a tradition of work-study projects with nearby Immunex. Laboratory manager Carol Ranthum says that about 15 of the company's 140 staff members are students who work 10 to 20 hours a week; the state pays two-thirds of their salaries. Immunex has hired five graduates of the programme since 1982. In addition to saving money, work-study arrangements and adjunct professorships also serve to remedy the threat of stagnation that plagues any high-tech curriculum.

Ultimately, however, employers seem to value experience as highly as either two or four years of education. "The more skills you can bring in that don't have to be taught, the better off you are", says Pioneer Hybrid's biotechnology research director, Nick Frey. By combining experience with education, biotechnology degree programmes are bound to ensure that tomorrow's graduates have at least a little more laboratory finesse than today's.

Karen Wright

Climatic research

Wind of change in Washington

Washington

AN unlikely alliance between environmental activists and former government officials has been formed to warn decision-makers and the general public about the serious climatic changes that many scientists now predict could occur within the next 100 years. The Climate Research Institute, based in Washington DC, plans to publish a newsletter giving non-technical summaries of research results and to organize conferences on the effects of climate change for constituencies such as coastal communities and agriculture.

The institute says there is a "solid scientific consensus" that a perceptible global warming due to an enhanced greenhouse effect will soon be apparent, and that an

Is that my brokers? Buy into Air Conditioning...



increase of 5°C is a reasonable prospect by the middle of next century. The institute's president, John Topping, a lawyer who was formerly staff director of the Environmental Protection Agency's office of air and radiation, says the institute was created to draw public attention to such projections and to put them high on the public policy agenda. He points out that only 9 out of 12,000 employees of the Environmental Protection Agency are working full time on climate change.

The institute's board includes former officials from both the executive and the legislative branch, as well as luminaries from the environmental movement. It will sponsor discussion and, Topping hopes, some policy research both on means of delaying climate change and on means of mitigating its consequences. The institute may, for example, pay researchers' expenses to visit Washington to address government officials, and it hopes to organize national and international conferences on selected aspects of climate changes.

As yet, the institute cannot afford to pursue all these avenues, but Topping is hoping that current approaches to corporations and charitable foundations will reap rewards.

Tim Beardsley

US technology exports

That won't do nicely at all

Washington

LAUNCHED with a flourish earlier this summer, a scheme by the Department of Defense and the Department of Commerce to remove some of the red tape bedevilling exporters of high-technology products has failed to interest the industries it was designed to help. Originally called the "gold card" scheme, the proposed regulation creates a list of certified end-users (CEUs) to whom controlled products may be sent without a separate export licence for each shipment. Although the negative response has not prompted the Commerce Department to abandon the scheme, officials acknowledge that revision is required.

The official comment period on the proposed regulations ended on 22 August. As originally proposed, CEU status would be reserved for importers of US goods that demonstrated and maintained "an exceptional record of compliance with US export and re-export control regulations". CEU status would at first be limited to companies in countries belonging to the Coordinating Committee for Multilateral Export Controls (COCOM), plus Canada. When an exporter wished to ship to a CEU, he would simply call the Commerce Department and receive an export clearance over the telephone.

While some in industry endorse the idea of the CEU scheme, most feel it needs to be modified to become effective. Bill Maxwell of the Computer and Business Equipment Manufacturers Association says foreign companies will have to be given incentives to apply for CEU status.

The proposed regulations were designed mainly to help small to medium-sized exporters. Michael Ciesinski of the Semiconductor Equipment Materials Institute (SEMI), whose members are mostly just such companies, likes the scheme in theory. But for regulations to be useful, Ciesinski believes they must apply to a broader class of products, as well as a larger number of countries. SEMI members are also unhappy that distributors are not included as potential CEU recipients, as many small to medium-sized company export their products through distributors rather than direct.

Some in industry do not even like the concept of CEUs. Edward Black, vice-president of the Computer and Communications Industry Association (CCIA) feels that the CEU proposal actually adds more regulatory hurdles than it removes. He notes that foreign companies would be subject to audits to make sure that they were maintaining their "exceptional record of compliance".

A regulation proposed by the Commerce Department to loosen controls on

re-export of products containing components made in the United States has received a much warmer welcome from industry. The new proposals would remove controls on products containing less than 20 per cent US parts. Boyd McKelvin, head of the Industry Coalition on Technology Transfer, calls this proposal a "major step", but adds that it too should be applied to more countries than those included in the proposed regulations.

Congress has its own agenda for relieving the regulatory burden on US exporters. An omnibus trade bill passed by the House of Representatives would remove 40 per cent of the items now restricted by Export Administration regulations. Some

in the House feel that the CEU scheme is an attempt to mollify members who feel the Department of Defense should not have so great a role in deciding what commodities should be restricted. The original idea for CEU came from Stephen Bryen, deputy Under Secretary of Defense and director of the Defense Technology Security Administration. The Senate may take up its own version of a trade bill this fall, but a comprehensive revision of the trade laws seems unlikely to be approved by Congress this year.

Implementation of the CEU scheme is not imminent. David Schlechty of the Office of Export Administration says comments on the proposal will definitely send Commerce officials "back to the drawing board". But CCIA's Black says they should "take the drawing board and throw it away".

Joseph Palca

Particle accelerators

Synchrotron fever grips Japan

Tokyo

SYNCHROTRONS are all the rage in Japan, with government institutes, universities and industry all eager to have sources of their own. The Science and Technology Agency (STA) has now weighed in with a proposal to build a giant 6 GeV synchrotron by 1994 that would place Japan at the forefront of synchrotron technology. But experts wonder whether the agency can really do it.

Japan already has several synchrotrons in the range of 0.4 to 0.7 GeV and one giant: the 2.5 GeV "Photon Factory" at Tsukuba. Nippon Telegraph and Telephone Corporation (NTT) has built one beam line and three stations at the Photon Factory at its own expense, and Hitachi, NEC and Fujitsu are soon expected to come on-line with their own beams. NTT also intends to construct its own synchrotron at its Atsugi laboratory, while one enterprising Japanese company is planning to market compact synchrotrons with superconducting magnets. In the words of Professor Kazutake Kohra, former director of the Photon Factory, "synchrotron fever seems to have struck Japan".

The industrial users are mainly interested in using X-rays for VLSI lithography and for studies of semiconductor lattice structure, purposes for which synchrotrons in the 1 to 3 GeV class are adequate. But for production of hard X-rays and fundamental studies at the sub-ångstrom level more powerful electron accelerators of 5 to 10 GeV are necessary, which is where the STA plan comes in.

In 1983 a group of professors at Osaka and Kyoto Universities, feeling left out of the synchrotron race in the Kanto area around Tokyo, proposed that a 6 GeV synchrotron be built in the Kansai area at Nishi Harima "technopolis", a new high-

tech town in Hyogo Prefecture. STA was the only government organization that seemed interested.

The provisional plan calls for construction of a 400 m 2 GeV linear accelerator to feed a synchrotron with a 300-m diameter storage ring and 30 to 40 beam lines. 100 million yen (£0.4 million) has been requested from the Ministry of Finance to set up a design committee. Construction would start in 1989 and total cost is estimated at over ¥60,000 million (£260 million). STA also has an ambitious plan to add a free-electron laser to the facilities, which would boost the bill by about another ¥20,000 million.

A key justification for the project is a familiar one: until recently Japan has been "catching up on the West" but now it is her turn to take a leading role in fundamental research. But can STA do it?

Nearly all of the machine physicists with expertise in building such a large accelerator facility are based at the High-Energy Physics Laboratory in Tsukuba. Participation of these scientists in this project would seem essential and STA is hoping for their cooperation. But the Ministry of Education, Science and Culture (MESC), which supports them, has its own plans.

Alongside the Photon Factory at Tsukuba is the 6-8 GeV accumulator ring of the TRISTAN electron-positron collider, and next year MESC is expected to come up with the ¥1,000 to 2,000 million to add an undulator that could provide one or two beam lines of hard X-ray synchrotron radiation. In addition, MESC has a plan to build a 2 GeV synchrotron either in Kansai or Hiroshima. If all these plans go ahead the synchrotron experts at the High-Energy Physics Laboratory are going to have their hands full, even without the STA project.

David Swinbanks

British Association

Science festival fun and furore

NORMALLY as British as cricket and cream tea (to be consumed after a charabanc ride to a site of local scientific interest) this year's meeting of the British Association for the Advancement of Science at Bristol was electrified by the fierce blasts of controversy. First among the outspoken was undoubtedly Mrs Wendy Savage, just cleared of charges of professional incompetence arising from her work as an obstetrician, and ready to put men in the dock for trying to take away control over childbirth from women.

To a crowded lecture theatre Savage quoted some startling statistics. In the United States the caesarian birth rate has risen to 20 per cent from 5 per cent in a little over ten years; Britain has just reached 12 per cent. But in neither country has there been commensurate improvements in perinatal survival rates. Savage cites numerous other cases of needless medical intervention: routine ultrasound scans and ever-rising birth induction and forceps delivery rates. Whereas in 1960, 30 per cent of British babies were born at home, now women have "lost confidence in themselves and in their ability to give birth". Mrs Savage most of all blames male obstetricians' desire to control women for the "medicalization of birth".

The main themes of the conference, appropriate for the year, were industry; the powers and limitation of science and technology; and computers and intelligence. Inevitably, industry meant how research, particularly basic research, could better serve industry. But new ideas were



The world's first scientific street performer? Dr Francis Evans of Sheffield Polytechnic was out every day during the British Association meeting explaining the principles of water power (and getting very wet) on a street corner. He claims to have seen one young "entrepreneur" passing the hat around among the crowds he attracted.

few and far between: listeners were launched onto a veritable ocean of platitudes as speakers listed the endless benefits of basic research, the peculiar aptitude of the British for it, and the shameful irresponsibility of the government in failing to give the support it requires. What of the great British inventions that were commercial flops? According to Sir George Porter, president of the British Association and the Royal Society, the universities are now enthusiastic about industry and "the days of ivory tower snobbishness about 'pure' research are long since past".

It took Professor John Ashworth, vice-chancellor of the University of Salford, to let in fresh air with the iconoclastic suggestion that Britain's vaunted record in winning Nobel prizes might really be a symp-

tom of weakness. His argument is that universities have, through the "peer-group pressure of the academic establishment as transmitted through the University Grants Committee" increasingly concentrated on the production of those "fitted for one very specific kind of vocation — that of research scholar". The latest moves by UGC to grade universities by research results are but the culmination of this process. It is assumed, he says, that those who fail to become scholars will find another vocation (in industry?) for which their scholarly training will stand them in good stead. He put his question bluntly "do failed Nobel laureates make the best recruits to industry?"

Ashworth's opinion is that industry needs a more highly educated workforce at all levels. His radical solution is to increase student numbers by a third, and to replace the three-year honours degree with a two-year general degree. A further two years' honours course would be provided only for those who need to specialize.

Elsewhere the British Association meeting continued in its own way: a curious pot-pourri of informing the general public of science (while avoiding being too outspoken), and involving the young in science (without attacking the education system) and allowing all, very politely, to have their say at question time (even the garrulous eccentrics). Two thousand of the meeting's three thousand attendees were young people, brought in by the sterling work being done by the British Association for Young Scientists, entertained by boffins from every conceivable discipline (more than 200 lectures and events in five days) and if they had a spare moment able to choose among a "feminist history walk", a tour of country pubs and a geological visit to the Mendips, to name but a few.

Alun Anderson

Murder charge for anthropology student

Washington

THE graduate student who found anthropologist Dian Fossey murdered in her mountain cabin last December is now the target of an international warrant for arrest issued by Rwanda's minister of justice. Wayne McGuire, a University of Oklahoma doctoral candidate, fled the African country last month.

McGuire came to the camp on Mount Visoke in mid-1985 with the intention of spending two years studying the dwindling population of mountain gorillas whose plight Fossey had fought to publicize. He had worked with Fossey for only five months when she was killed by a machete blow to the head. Fossey's vigilante tactics in nabbing poachers who attacked her animals won her many enemies; initial speculation as to her murderer's identity focused outside her own camp. McGuire proceeded with his work until late July when,

according to McGuire, a consular official hiked to the camp and advised him that he should seek legal advice.

McGuire claims Rwanda's charges were trumped up to avert negative publicity. He suggests the Rwandans, who received over \$20 million in US economic aid last year, fear that reports of Rwandan citizens killing well-known Americans would leave a bad taste in the mouths of US budget committees. The Rwandans are sticking to their guns: but, lacking an extradition treaty with the United States, their guns are not exactly loaded.

The University of Oklahoma has offered McGuire a teaching position starting this spring, but without access to the gorillas, his dissertation is scuttled. McGuire's attorney has written to the State Department requesting its assistance in having the warrant recalled so McGuire can return to Mount Visoke.

Karen Wright

Indian science

Foreign software giants pounce

New Delhi

INDIA now has a new kind of brain drain to worry about. Added to the old problem of the loss of highly qualified scientists to posts abroad is an "internal haemorrhage" to foreign high-technology companies that are coming to India to take advantage of the low salaries.

Liberal policies on collaborations allow drugs and electronics companies to switch some labour-intensive development work, such as software engineering, to India, where salaries are a third to a fifth of those in the West. The major losers are the national laboratories and universities, which see some of their brightest scientists going to the new arrivals.

Texas Instruments of Dallas, Texas, has already recruited 25 top Indian software engineers for its wholly owned subsidiary at Bangalore, India's Silicon Valley. Texas Instruments will use the programs developed to design very large-scale integrated (VLSI) chips. The programs will be exported via satellite to Dallas using the company's own Earth station at Bangalore. On labour costs alone, Texas Instruments expects quickly to recoup its \$6 million investment.

Following hard on the heels of Texas Instruments is Peat, Marwick and Mitchell, the world's largest accountancy firm, which is planning to set up an even bigger facility near New Delhi to prepare software for its worldwide computer operations. And the US company Citibank has set up a subsidiary near Bombay Airport to produce software for Citibank's internal use and for sale worldwide.

Last year, California's Fortune Systems signed a deal to buy software developed by Kirkoskar Computer Services in Bangalore for worldwide distribution. Other electronics companies such as Digital Equipment Corporation and

Hewlett-Packard have hired Hinditron of Bombay to develop new programs.

The Japanese are also cashing in on cheap labour: last year PSI Systems was established in Bangalore to develop software for Japanese companies. The Indian Department of Electronics, which approved these collaborations, is happy that the software deals are earning foreign exchange while keeping the best brains at home and gainfully employed.

But it is not just the software engineers who are in demand. The Swedish pharmaceuticals company Astra has set up a subsidiary in India to carry out basic research on drugs for tropical diseases. The company will be able to use actual pathogens (instead of test strains) and to conduct animal trials in India at one-twentieth of the cost in Sweden. More important, the subsidiary can test the finished drug on Indians before Astra markets it in the developing world. Astra's subsidiary in Bangalore has taken the cream of Indian biochemists, including the chief biotechnologist of the Indian Institute of Science. The Astra project was strongly opposed by the Council of Scientific and Industrial Research (CSIR) in that it drained Indian talent for the development of drugs "that will ultimately be sold to us". The Department of Biotechnology, however, considers it an "excellent model" of collaboration.

Astra's successful entry has encouraged other drug multinationals. Pharmacia and Hoechst are already setting up research centres in India. At least six other drug companies have submitted similar proposals to the Indian government. Critics say the invasion by foreign subsidiaries is a form of technological colonization, and that they are taking away India's best brains to fight its own companies.

K.S. Jayaraman

Hungarian science

Travel broadens the budget

Budapest

HUNGARY's new system of grants for basic research comes into operation this month. An injection of 4,000 million forints (£66 million) for the period to 1990 makes Hungary one of the few countries in the world, and the only one in the socialist bloc, to be increasing its investment in pure research. But the sums involved are relatively modest even by Hungarian standards (a senior university professor earns about 15,000 forints a month basic) and it is not yet clear how much of the available funds will be in "hard currency".

The new fund comprises 2,000 million forints for instruments and equipment, to be shared out among research institutions and universities, and 2,000 million forints to be distributed for projects carried out by individual scientists or teams. Project grants have been decided by two special committees of the Academy of Sciences (for the natural sciences and for the social sciences and humanities).

The need for the grants was explained last year by the General Secretary of the Academy of Science, Dr Lenart Pal, who admitted that without a rapid injection of cash Hungarian science would lose its ability to make major advances at the international level.

The lack of money for pure research has had several results. More scientists have taken on contract work for industry and agriculture. And, because hard currency is in short supply, scientists have to apply, often months in advance, for special permission to purchase standard reagents from Western countries, while many major items cannot be bought at all.

Foreign travel is fast becoming a mainstay of Hungarian science. This too costs hard currency — but far less than providing facilities in Hungary, and much can be accomplished by exchanges with foreign laboratories. But this can mean that researchers have to plan experiments for their next trip abroad and spend their time in Hungary writing up their previous trip. This promotes international cooperation, but is not particularly beneficial to young researchers.

Finally, there is the problem of keeping abreast of world literature. So far, most Hungarian institutions are managing to keep up their hard currency subscriptions, though constantly rising subscription rates mean that every year one or two of the less frequently used titles have to be dropped.

The new research grants are not intended as an advance for Hungarian pure research — rather as a holding operation to stop, or at least retard, a downhill slide.

Vera Rich

Small step forward for ice minus bacteria

Washington

ADVANCED Genetic Sciences fights on. The California biotechnology company, fined earlier this year for improper testing of its ice-minus bacteria, has announced Environmental Protection Agency (EPA) approval of its second round of test results. This frees the company to find a site for field trials but does not revoke the suspension of its experimental use permit, imposed in March, which cannot be lifted until the new site is evaluated. Delay has dogged ice-minus experiments since Advanced Genetic Sciences first obtained the go-ahead in November 1985. Tests of the *Pseudomonas* bacteria, which protect plants against frost damage by thwarting

ice nucleation, have been the subject of federal lawsuits and congressional subcommittee hearings (*Nature* 320, 2; 1986). Similar tests proposed by the University of California have weathered environmental impact assessments and a temporary restraining order. But the university's recent pledge to re-examine safety issues will postpone its experiments until at least next spring.

Local opposition and EPA requirements forced Advanced Genetic Sciences to abandon its site in Monterey County. But a company representative said that since last week's announcement, they had received many calls from residents of other areas offering test sites.

Karen Wright

Harvard anniversary

Oldest US college looks to future

Cambridge, Massachusetts

"By Universe, I mean the universe in which Harvard is located." Thus did Professor George B. Field, professor of astronomy at Harvard University, ironically echo the mood of three days of celebrations of the 350th anniversary of the oldest university in the United States, some thousands of whose 230,000 living alumni had come back hoping to be told that Harvard remains central to everything.

The celebrations have been at once a spellbinding demonstration of the richness of Harvard's scholarship and a proof that, whatever the intent, sentimental banality is unavoidable on these occasions. The hard core of the programme has been a series of a hundred or so symposia in which academics have given thumbnail accounts of their work, mostly both witty and articulate. The schmaltz is typified by the closing ceremony at which Mr Walter Cronkite, the legendary TV anchorman with no previous Harvard connections, read a script explaining that after sustaining Puritanism for more than a century and then driving off the British, Harvard alumni went on to banish the great depression, to win the Second World War, to solve the Cuban missile crisis and to launch the drive into worlds beyond our World. The alumni, having come to hear just this, went off into the night humming "Fair Harvard....".

The speakers' rostrum has as often been occupied by the Great and the Good, bearing greetings from elsewhere, as by the presidents of the network of alumni associations (in which, as one put it, membership entails "no dues, but from which there is no escape"). President Ronald Reagan was nevertheless conspicuous by his absence; he had been invited, as is traditional, but declined only after Harvard had decided to break with tradition and not to award its distinguished speakers honorary degrees.

The Prince of Wales, heir to the British throne and pinch-hitting for his father, who is the Chancellor of the University of Cambridge and Harvard's traditional guest on these occasions, made a nicely judged plea that "we have not only to teach men how to make things, but also to produce people who have complete moral control over the things they make". And, to come to grips with the "dark side of man's psyche", he urged that more attention should be paid to the "natural science of psychology", a cause to which the British royal family has hitherto paid little attention.

Mr George Schultz, the US Secretary of State, arguing that it takes a free country to develop better means of communication and thus an information revolution,

attacked protectionism (and the US Congress for harbouring it), moralism (meaning attempts to force the pace of change in South Africa) and isolationism (particularly the Congress's attack on the State Department's own budget). Mr Shultz's rehearsal of the US administration's present international stance was spiced with a protest at the arrest in Moscow on a charge of espionage of Mr Nicholas Daniloff, a Harvard graduate of 1956.

Mr Caspar Weinberger, US Secretary of Defense, made a similar protest at a symposium on the Kennedy School of Government's project under the title "avoiding nuclear war", but then went on to disappoint an expert audience because his restatement of the administration's view that the Strategic Defense Initiative is the "better way" than mutually assured deterrence lacked the subtlety expected on an academic occasion. But Mr Weinberger, fair play, was having to deal with notes and whispers about the hijacking of



Me? Yale, I'm afraid...

the Pan-American aircraft in Pakistan. As at Sheik Yamani's earlier breathtaking plea for stable oil prices, there were armed Boston policemen (and men with bulging jackets and pentagonal badges) standing around the lecture room.

The symposia have nevertheless been extraordinarily stimulating. Field's symposium had Professor Steven Weinberg confessing himself a convert to superstring theories of elementary particles and Professor Irwin Shapiro saying that it is only a matter of time before very long baseline interferometry yields absolute distances to at least one quasar with an interposing gravitational lens. The parallel computer people showed that their field is full of promise yet still in its infancy. Professor Richard Wilson's moderate defence of nuclear power excited a spate of moderate questions. Molecular biology, of which Harvard has cause to be proud, was curiously muted, with Professor Walter Gilbert's account of what things are possible unsupported by the medical school, which has had almost nothing to say for itself.

Much talk has been about the future,

that of the world and, inevitably, of Harvard. If money is all that matters, Harvard must be among the safest of institutions. By the end of June last year, its endowment funds had topped \$3,000 million, and contributed \$110 million a year of the total operating costs of \$650 million a year. University administrators elsewhere are especially envious of the way the university finances new projects by issuing tax-exempt bonds and of the deliberate way in which maintenance is planned by the yardstick of the replacement cost of buildings. The average professor's salary in the Faculty of Arts and Sciences now amounts to \$66,000, the highest in the United States (and nearly twice that of people without tenure). The endowment fund explains why 85 per cent of undergraduates (among whom American-Asians now amount to 9 per cent) receive some form of scholarship aid, which takes the edge off some academics' shame that the cost of a Harvard education should now be \$17,000 a year.

The cost is one threat to the future and to open access. Dr Derek Bok, president of Harvard for the past 17 years and a low-key orator with a reputation for compromise, identified another last week: "as people read of the monetary value of a Harvard diploma, the swelling size of the endowment and the fabled influence of the old-boy network, fascination and respect easily turn to envy and resentment".

But Bok's main message was embedded in a plea for functional independence from government and freedom from pressures to help solve social problems in ways that compromise the proper functions of the university (which was generally understood to refer to the way anti-apartheid activists prevented 350 university guests from dining on Thursday evening). It is the danger that the growth of extracurricular activities by professors will create a generation of academics seeking "to combine the freedom and security of a tenured academic post with the income and visibility traditionally reserved for people who take greater risks".

John Maddox

Medals for twenty

INSTEAD of the usual honorary degrees, Harvard this year awarded university medals to twenty people including Professor Margaret Kivelson (Earth and Space Sciences, University of California, Los Angeles); Dr Adetokundo O. Lucas (now with the Carnegie Corporation, but director of the WHO programme for research training in tropical diseases until earlier this year); Dr Raymond J. Nagle (now retired as professor of Dental Medicine at Harvard); Dr Edward Purcell (retired professor of physics at Harvard) and Dr Carl W. Walter (professor of surgery until his retirement in the early 1960s). □

Biotechnology guidelines

SIR—Tim Beardsley's article on the recently issued US biotechnology guidelines (*Nature* 322, 398; 1986) indirectly implies that this association has expressed the belief that the exemption of well-defined non-coding regulatory sequences from the administration's regulatory proposal should be removed. This is not our position. The Industrial Biotechnology Association (IBA) testified at the congressional hearing that Beardsley reported on and our statement supported the exemption contained in the proposed policy for biotechnology regulation.

We believe that many people have misinterpreted the nature and extent of the exemption for non-coding regulatory sequences. We have neither in public nor in private advocated its removal from the regulatory policy statement. IBA believes that a good case can be made for its inclusion and intends to address the issue in comments to the regulatory agencies.

RICHARD D. GODOWN
*Industrial Biotechnology Association,
2115 East Jefferson Street,
Rockville, Maryland 20852, USA*

SIR—The article by Tim Beardsley on the biotechnology guidelines (*Nature* 322, 398; 1986) inaccurately implies that the Association of Biotechnology Companies (ABC) wants the exemption of well-defined non-coding regulatory sequences removed from the US regulatory agencies' biotechnology proposal, published on 26 June 1986. On the contrary, ABC believes that removal of the exclusions would be scientifically indefensible. The concepts embodied in the definition for this exemption are derived from years of evolution of the National Institutes of Health (NIH) Recombinant DNA Guidelines. Those guidelines recognized, on the basis of credible scientific data, that the splicing of DNAs that were "well characterized and free from harmful sequences" could be exempted from the guidelines, or at least that the required level of physical containment could be reduced.

The Association of Biotechnology Companies, representing 165 members worldwide, believes that the proposed biotechnology regulatory policies will evolve, as the NIH guidelines have, into less stringent ones as more understanding and knowledge is obtained. The removal of the exclusions, as your article proposes, would create an avalanche of unnecessary paperwork for government and industry and cannot be justified scientifically.

The article therefore did not accurately represent ABC's views, nor, I believe, those of the Industrial Biotechnology Association.

The headline "Regulations please nobody", taken out of context, also raises

concern by conveying an overly negative impression. ABC supports the publication of the US biotechnology regulatory proposals. We recognize that it will be an evolving document, as were the NIH guidelines. Those areas, language, definitions or rules in the guidelines that ABC finds inappropriately burdensome can be ameliorated through comments, testimony and encouraging academic investigators to speak out. ABC also recognizes that no document will be entirely non-controversial, given Jeremy Rifkin's strategy of appealing to the emotionality of the biotechnology issues.

The Association of Biotechnology Companies calls upon all interested parties to establish a constructive dialogue rather than seek to polarize interests. For the benefits of biotechnology to be achieved with minimal risks, a partnership between the public, the government and industry must be developed to cut through the emotionality, misconceptions and negativity that have been raised about biotechnology debate.

BRUCE F. MACKLER
*Association of Biotechnology Companies,
1220 L Street NW, Suite 615,
Washington, DC 20005, USA*

• The false impression that industry organizations are critical of the exemption from special review for noncoding sequences is due to an ambiguity introduced in the editing. In the form of words originally used by Tim Beardsley, it was clear that the critical opinions are held privately by some industry spokesmen.— Editor, *Nature*.

AIDS in Africa

SIR—Concern about the effects of the epidemic of acquired immune deficiency syndrome (AIDS) has related predominantly to the developed countries, but AIDS is a disease affecting people everywhere and in developing countries, such as those of central Africa, coping with AIDS presents particularly difficult problems.

In the developed world, extra money from government can be grafted onto a health service with an established infrastructure, which operates alongside charities, a sophisticated research capacity and an efficient local government. Furthermore, in such an environment, government can legislate about aspects of management and can gather essentially confidential information and use it to advise on therapy.

In developing countries, there is a very different social and economic environment. Many of the hospitals still run short of essential drugs and diagnostic reagents, and disposable needles and syringes are

commonly re-used. Infant mortality remains at 10 per cent in the first year mainly through preventable causes such as measles and diarrhoea. Even communication can be a problem because of a shortage of paper.

If resources are so stretched now, it is logical to assume that even the slightest reduction in the money available for the provision of basic health could have the most serious of consequences. AIDS stands to deplete and divert resources vital for basic health services. Health departments could, of course, argue that AIDS is incurable and refuse to spend money on prolonging patients' lives. It is difficult to know how much might be saved but it should be borne in mind that the average AIDS patient costs the British National Health Service £10,000. But even if this policy were to be considered, AIDS would still divert large sums of money because there would remain a need to confine the disease: it would cost several million pounds to mount public information campaigns, to provide disposable needles and syringes and to monitor blood donors.

A rigid policy of containment rather than treatment is, in any case, unworkable. If the disease spreads, as it has in parts of Uganda where 2,500 people have already died and an estimated 30 per cent of the population are affected, then health workers dealing with the patients will be unable to withhold help. Whatever policy is adopted, it is impossible to know at the beginning of treatment whether a sickly child has been infected by an AIDS-positive mother, or whether a patient who develops malaria or tuberculosis does so because of impaired immunity.

The equations that have been used for determining health economics will now need to include an AIDS factor. That factor might be relatively small in the developed world, but elsewhere it may prove a drain on resources so large that basic health services are undermined.

There is, however, little prospect that the extra resources needed can come from the developing countries, so if the richer countries do not provide them, nobody will. Developed countries cannot stand by and allow AIDS to devastate, and at worst destabilize, central Africa, and later parts of South America and possibly Asia. There is an urgent and compelling need for the governments and charities of the developed world to provide support (and this will have to be vast) for the developing countries, both to underpin conventional health provision and to help to contain this treacherous epidemic. There must be no delay.

JOE COLLIER
*Department of Clinical Pharmacology,
St George's Hospital Medical School,
Cranmer Terrace,
London SW17 0RE, UK*

Joe Collier is a World Health Organization Consultant on the Essential Drugs Programme in Tanzania.

The shortest period binary star?

The discovery of a binary star with a period of only 11 minutes was announced last week at the 9th European meeting of the International Astronomical Union. What does it mean?

PERIODICITIES in astronomical X-ray sources are legion, but occasionally one is discovered that leads to the postulation of either a new class of astronomical object, or a familiar object in a remarkable configuration. The latter would seem to be the case with the discovery, announced last week, of a bright X-ray source showing a coherent periodicity of 685 seconds. The interpretation is that the source is a neutron star in an orbit of diameter less than a seventh the radius of the Sun about a low-mass white dwarf, making the system, if the interpretation is confirmed, by far the shortest period binary system known.

We owe these results and interpretation to W. Friedhorsky of the Los Alamos National Laboratory, and L. Stella and N.E. White of the EXOSAT Observatory, who presented them at the International Astronomical Union meeting at the University of Leicester last week.

According to these authors, the 685-second oscillations on the emissions of X-ray source 4U1820-30 (which lies at the core of globular cluster NGC6624) have an amplitude of only 3 per cent, but are clearly present in all four observations made by EXOSAT in 1984-85. Normally one would expect a period of this order to represent the spin of a magnetized neutron star accreting matter from a bright, massive companion star onto its off-axis magnetic poles, the rotation of the poles causing variations in the radiation received on Earth.

4U1820-30 is, however, an X-ray burst source, that is, at irregular intervals it flares dramatically with rise times of a few seconds, and none of the 30 or so X-ray bursters that have been studied shows coherent pulsations. If the periodicity in 4U1820-30 were associated with the rotation of a neutron star, it would violate the empirical rule that bursting sources do not pulse, and pulsing sources do not burst. The bursts are probably the result of runaway nuclear burning of accreted matter accumulated on the surface of the neutron star, and the empirical rule suggests that this process works only if the neutron star does not have a strong magnetic field, as is required to produce regular pulses of X-ray emission at the spin period of the neutron star.

Priedhorsky *et al.* did indeed find direct evidence that the 685-second periodicity is unlikely to be the rotation of the neutron star. Since a neutron star generally gains

angular momentum from the material that it accretes, one expects to see the spin period decrease with time. For a magnetized neutron star the lever arm of the accreting matter is large, and the moment of inertia relatively small, and as a result the spin-up can in many cases be measured. For 4U1820-30 there was no evidence for any spin-up in the year-and-a-half spanned by the EXOSAT observations, and the *post facto* discovery of the same period in observations made by the SAS-3 satellite in 1976 placed even stricter limits on the rate of any possible period change. Priedhorsky *et al.* argue that, given that spin periods are ruled out, this stability is characteristic of an orbital period. For such an ultra-short period the two stars must be separated by only about one seventh of the radius of the Sun, and the unseen companion to 4U1820-30 must be unusually small: the most likely candidate is a very low mass (about 0.07-solar-mass) white dwarf.

The main difficulty in understanding this result, however, is to see how such a binary system could have formed. The evolution of 'normal' low-mass binary systems in which a neutron star accretes from a small hydrogen-burning companion star never proceeds to periods shorter than 80 minutes, since the companion becomes too light to burn its nuclear fuel and starts to expand in response to mass loss. Fortunately there are other possibilities for 4U1820-30, which lies at the core of a globular cluster, one of the dense aggregations of some hundreds of thousands of stars containing some of the oldest stars in the galaxy.

It has long been known that X-ray binaries are much more common in globular clusters than one would expect from the numbers found elsewhere in the Galaxy. Accordingly, one needs a formation mechanism special to globular clusters. The very high stellar densities in cluster cores suggest that this mechanism may be the gravitational capture of companions by neutron stars. In another contribution to the meeting, F. Verbunt (Max-Planck Institut für Astrophysik, Garching) showed that capture of a low-mass red giant star by this process can be reasonably common. The core of such a star consists of degenerate helium (that is, it is an incipient white dwarf) surrounded by a large, tenuous envelope containing, initially at least, most of the star's mass. About one third of the neutron star-red

giant captures are expected to be direct collisions, leading to the neutron star orbiting inside the giant's envelope. The resulting intense frictional drag on the neutron star causes it to lose angular momentum and hence to spiral inwards towards the degenerate core. At the same time the envelope is heated and is ultimately ejected entirely. Thus, the result of this collision and spiral-in process is the formation of a very close binary system consisting of a neutron star and a helium white dwarf. Depending on the stage of evolution of the giant when the capture occurred, the white dwarf can have a mass anywhere between 0.075 and 0.45 solar masses.

The subsequent evolution of such a binary is easy to predict: loss of angular momentum by gravitational radiation will cause the stars to spiral closer together and the period to shorten. Eventually at a period of a few minutes (depending on the white-dwarf mass) the stars will be so close that matter can be pulled off the white dwarf, leading to sufficient accretion onto the neutron star to turn it on as a bright X-ray source. As Verbunt showed, the properties of 4U1820-30 are generally as expected if it is an ultra-short period binary of this type.

Although the interpretation of this startling new result in these terms seems very appealing, the amplitude of the 11-minute periodicity is small, and one can think of other ways of producing it. For example, as suggested by Trumper (Max-Planck Institut für Extraterrestrische Physik, Garching) in the discussion following these papers, the period could represent the precession of a fast-rotating neutron star. Another possibility is that the periodicity is actually associated with an entirely separate (and presently unknown) X-ray source which just happens to lie within the field of view of the EXOSAT detectors when observing 4U1820-30. Because of their low formation rate and short life-times, systems like 4U1820-30 are expected to be rare. It is rather ironic, but fortunate for us, that it is the brightest globular cluster X-ray source, and the only one in which such a small amplitude periodic signal could have been detected with present instruments.

A.R. King & M.G. Watson

A.R. King is in the Astronomy Department and M.G. Watson is in the X-ray Astronomy Group, Physics Department, University of Leicester, Leicester LE1 7RH, UK.

Animal navigation

The bee's celestial compass

from R. C. Hardie

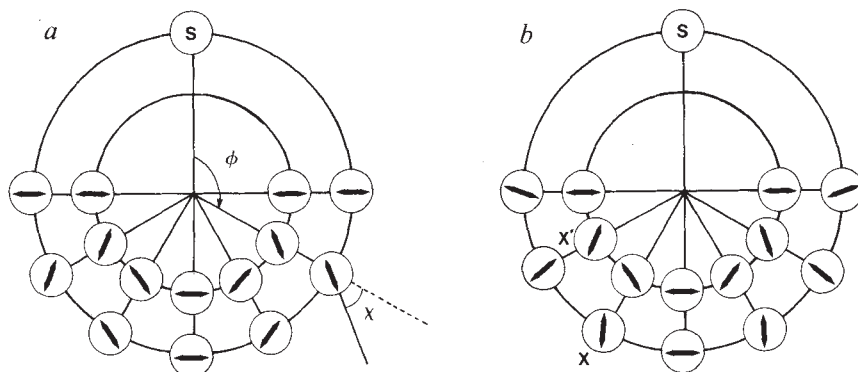
Forty years ago, the famous experiments of Karl von Frisch showed that bees can navigate using polarized light patterns in the sky¹. The mechanism by which they perform this feat remained unknown until a recent series of experiments led Rossel and colleagues to propose a simple, comprehensive solution to this classic problem²⁻⁴. On page 128 of this issue⁴, Rossel and Wehner describe an ingenious and stringent experiment to test their hypothesis, which emerges unscathed.

The polarized light pattern in the sky derives from scattered sunlight, the direction of polarization (*e*-vector direction) always being at right angles to the plane defined by the observer, the Sun and the point in the sky being observed. A predictable but rather complex pattern is thus generated that changes as the Sun moves through the sky. The bee's problem is to deduce the Sun's bearing from the *e*-vector pattern even when only one small patch of blue sky is visible. The essence of the hypothesis of Rossel *et al.* is that, rather than performing celestial trigonometry or being informed of the patterns in all their detail, the bee possesses only a simplified map of the *e*-vector distribution in the sky.

In the bee's map (see figure) any given *e*-vector direction is assumed to have a fixed bearing with respect to the Sun irrespective of elevation in the sky or time of day. For example, a patch of sky with a horizontal *e*-vector lies exactly opposite the Sun (180°), a vertical *e*-vector at 90° to the Sun and so on. Ambiguities arising from the fact that any *e*-vector direction, in general, occurs at two positions in the sky are avoided by assuming that the *e*-vector lies in the half of the sky opposite to the Sun⁵.

The map is embodied in an array of ultraviolet-sensitive photoreceptors in the dorsal rim of the eye⁶. Each photoreceptor is maximally sensitive to a particular *e*-vector direction, the direction rotating from the back to the front of the eye⁷. To use this map Rossel and Wehner suggest that the bee simply turns until the retinal map is in register with the *e*-vectors in the sky. When this is achieved, the photoreceptors generate a maximal signal that tells the bee that she is pointing directly away from the Sun. Because the bee's map is invariant, and corresponds only to the actual *e*-vector distribution at dawn and dusk, mistakes are bound to occur under certain conditions. It is these mistakes that led Rossel *et al.*² to deduce the bee's generalized map.

The mistakes, central to the hypothesis of Rossel *et al.*, distinguish it from the two other major hypotheses that seek to explain the bee's performance: (1) that the bee uses some abstract knowledge of the laws of scattering to reconstruct the Sun's position by celestial trigonometry⁸; or (2) she has a precise knowledge of the real *e*-vector distribution in the sky⁹, either from a memorized 'snapshot' or from a hard-



The bee's map (inner circle). A given *e*-vector direction, χ , is assumed to lie on a fixed bearing, ϕ , with respect to the sun (S). The bee rotates until her retinal map is in register with the *e*-vectors in the sky (outer circle) and then assumes she is facing directly away from the Sun. At dawn and dusk (a), the celestial *e*-vector distribution matches the bee's map perfectly; with the Sun 24° above the horizon (b), a different *e*-vector pattern arises. If the bee can see the whole sky she will still orient correctly because the errors in the left and right halves are in the opposite direction and will cancel each other out. If, however, she can see only one point, for example X, she would rotate until X' coincides with X, thus making an error of 30°. (Adapted from ref. 3.)

wired celestial almanac. There is still a discrepancy in the literature because previous studies^{3,10} failed to detect mistakes like those made by the bees of Rossel and colleagues. Although no recent data contradict the extensive results of Rossel *et al.*, independent confirmation would be welcome.

In their early work, Rossel and colleagues showed that the single, invariant map of the bee is used under all conditions tested. In their new work they test the specific formulation of their hypothesis: namely that the bee rotates, scanning the sky with the map embodied in her retinal array and then interprets the direction which gives maximum output as pointing directly away from the Sun. With the real sky this occurs when the *e*-vectors in the sky are in register with the individual *e*-vector sensitivities of the photoreceptors. However, in the experiment the signal is generated by modulating the intensity of an ultraviolet light source using a feedback device to ensure that the intensity reaches a maximum when the bee reaches a predetermined position. The fact that this ingenious simulation does in fact result in the bee orienting just as she normally would do to the appropriate *e*-vector

provides direct evidence for the scanning strategy proposed.

Rossel and Wehner's hypothesis has all the hallmarks of being correct. It is a classically simple solution to one of the more impressive intellectual feats in the animal world. It has successfully withstood a battery of experiments, the results of which would be perplexing, to say the least, on any other hypothesis.

The bee's strategy, although capable of generating blatant navigational mistakes in experimental situations, is likely to perform quite satisfactorily under natural conditions. In particular, errors induced in the left and right halves of the eye will be opposite in sign and will thus tend to

cancel each other out. Further, as well as actually being correct at dawn and dusk, the bee's map of the celestial *e*-vector distribution also happens to represent the average values as they occur in the most polarized band of the sky³. In other words, with limited hardware the bee has developed the single map that is most likely to be correct at any time.

This is by no means the end of the story. Future developments are likely to occur on at least two fronts. First, it will be interesting to see how widespread this mechanism is. Many insects (including

1. von Frisch, K. *Experientia* **5**, 140 (1949).
2. Rossel, S., Wehner, R. & Lindauer, M. *J. comp. Physiol.* **125**, 1 (1978).
3. Wehner, R. & Rossel, S. *Fortsch. Zool. Bd.* **31**, 11 (1985).
4. Rossel, S. & Wehner, R. *Nature* **323**, 128 (1986).
5. Brines, M.L. & Gould, J.L. *Science* **206**, 571 (1979).
6. Labhart, T. *J. comp. Physiol.* **141**, 19 (1980).
7. Sommer, E.W. thesis, Univ. Zurich (1979).
8. Kirschfeld, K., Lindauer, M. & Martin, H.Z. *Naturforsch.* **30c**, 88 (1975).
9. Dyer, F.C. & Gould, J.L. *Am. Sci.* **71**, 587 (1983).
10. von Frisch, K. *The Dance Language and Orientation of Bees* (Harvard Univ. Press, 1967).
11. Labhart, T. *J. comp. Physiol.* **158**, 1 (1986).
12. Labhart, T., Hodel, B. & Valenzuela, I. *J. comp. Physiol.* **155**, 28 (1984).
13. Hardie, R.C.J. *J. comp. Physiol.* **154**, 157 (1984).
14. Fent, K. *J. comp. Physiol.* **158**, 145 (1986).
15. Strausfeld, N.J. & Wunderer, H. *Cell Tissue Res.* **243**, 163 (1985).

desert ants¹¹, crickets¹² and flies¹³ have recently been shown to have similar arrays of polarization-sensitive photoreceptors in dorsal eye regions, and the desert ant, which also navigates by celestial polarization patterns, makes analogous navigational errors to the bee¹⁴.

Second, progress may soon be made in unravelling the mechanisms of information processing. The work of Rossel and Wehner provides an indication of what

sort of interneurons we might expect to find, and some elegant anatomical studies in the fly, which use the powerful technique of trans-synaptic cobalt migration, identify many of the interneurons in the polarization-sensitive pathways of this insect¹⁵. □

R.C. Hardie is in the Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK.

Surface science

Helium atoms reveal phases

from J.A. Venables

HELIUM atom scattering is the ultimate surface tool, in that penetration into the bulk can be entirely neglected. The latest of several powerful tools for studies of surface structure and vibrations to reach maturity, the technique has in the hands of a group at Jülich, West Germany¹⁻⁴ revealed surface phonons on a platinum crystal and, in physisorbed xenon on platinum, two-dimensional phase changes and a series of discrete transitions towards bulk behaviour as the number of atomic layers is increased.

Such work clearly demonstrates that helium atom scattering is going to be important in unravelling the nature of 'two-dimensional physics' at surfaces and in adsorbed layers. But the technique has only recently become practicable through the development of high-pressure nozzle beam sources, effective differential pumping, high-energy resolution time-of-flight spectrometers and efficient single atom counting detectors. These non-trivial pieces of hardware, when attached to a large, but essentially standard, ultrahigh vacuum chamber, form the basis of the technique. The high-pressure nozzle gives a suitably monoenergetic incident beam with thermal energies of 0 to 30 meV, and the spectrometer can measure its energy width within 0.3 meV. The beams come into and leave the differentially pumped scattering chamber through small aper-

tures with overall angular resolution of around 0.3°; thus very detailed angular scans are possible to analyse the diffraction patterns, either by rotating the sample or the detector.

Because the beam has such a low energy, it is an easy matter to distinguish elastic from inelastic scattering. Elastic scattering consists of diffraction and diffuse scattering from defects, whereas inelastic scattering results from atomic vibrations (phonons), which cause energy changes in the range 0–20 meV. In this respect, the technique is similar to neutron scattering from bulk solids, which is the main technique for determining the energies and momenta of phonons. However, in sharp contrast to thermal neutrons which travel several centimetres through solids, thermal helium beams interact so strongly that they only probe the tail of the electron distribution of the outermost atoms. Consequently the helium beam does not penetrate into the bulk. Thus with improved intensities and energy and momentum resolution, helium scattering has become increasingly competitive with low-energy electron scattering for studies of surface vibrations, and with higher energy electron and X-ray scattering for studies of surface crystallography. Of course, these latter techniques still retain advantages; in particular X-ray scattering has the highest angular

resolution, whereas now electron techniques have the highest signal strengths.

These developments in helium atom techniques are well illustrated by the Jülich work on the surface phonons on a Pt(111) crystal¹, including the changes in phonon structure induced by oxygen chemisorption² and the structural and vibrational changes attendant on physisorbing xenon onto this same surface^{3,4}. The time-of-flight spectrum on the bare (111) surface is now sufficient to show up crystallographic anisotropies in the surface acoustic (Rayleigh) wave¹. The effect of chemisorbing oxygen in the p(2×2) structure is to halve the extent of the surface Brillouin zone, and to open up a gap at the zone boundary; this gap has now been observed very clearly for the first time². Theoretical calculations invoke particular force constants which, by adjustment to agree with experiment, give information on the strength of both central and non-central forces between Pt and O atoms at the surface.

The elastic diffraction peaks studied in the case of physisorbed xenon³ have been sufficiently sharp to study three separate phases: the commensurate (C) phase in which the Xe has the so-called ($\sqrt{3}\times\sqrt{3}$)R30° structure, in which the Xe surface mesh is $\sqrt{3}$ times the Pt mesh and is rotated by 30°; the incommensurate (I) phase compressed by around 6 per cent, and a further phase (R) compressed by around 10 per cent and rotated some 3°.

The corresponding C–I–R transitions have been seen previously for Xe adsorbed on graphite by both electron and X-ray diffraction; but this is the first time that helium atom diffraction has had sufficient angular resolution to see such transitions. Moreover strong scattering from the high atomic number bulk material would make Xe/Pt(111) difficult for either electron or X-ray diffraction. Because the helium beam sees only the outermost layer, the nature of the substrate is essentially irrelevant. In the latest work, the Jülich group have also shown that the time-of-flight spectra from these monolayer and multilayer structures are also very characteristic. Whereas the monolayer has an Einstein-like mode with essentially no dispersion, increasing layer thickness introduces dispersion, neatly revealing a discrete series of transitions towards the bulk phonon behaviour as the number of layers is increased⁴. □

1. Kern, K., David, R., Palmer, R.L., Comsa, G. & Rahman, T.S. *Phys. Rev.* **B33**, 4334 (1986).
2. Kern, K., David, R., Palmer, R.L., Comsa, G., Jin He & Rahman, T.S. *Phys. Rev. Lett.* **56**, 2064 (1986).
3. Kern, K., David, R., Palmer, R.L. & Comsa, G. *Phys. Rev. Lett.* **56**, 620 (1986).
4. Kern, K., David, R., Palmer, R.L. & Comsa, G. *Surface Sci.* (in the press).

J.A. Venables is at the School of Mathematical and Physical Sciences, University of Sussex, Brighton BN1 9QH, UK and the Department of Physics, Arizona State University, Tempe, Arizona 85287 USA.



Meteoritic nitrogen

A slice through the Bencubbin meteorite found in Australia shows the shock-welded matrix surrounding metal with silicate and chondritic clasts. On page 138 of this issue, I.A. Franchi, I.P. Wright & C.T. Pillinger of The Planetary Science Unit at the Open University, UK show that virtually all the nitrogen in the meteorite is enriched with ¹⁵N. The authors discuss the implications of these results for the origin of the components of the meteorite. The meteorite is about 65 cm wide.

Networks

Simple is the best for dynamic routing of telecommunications

from Alistair Mees

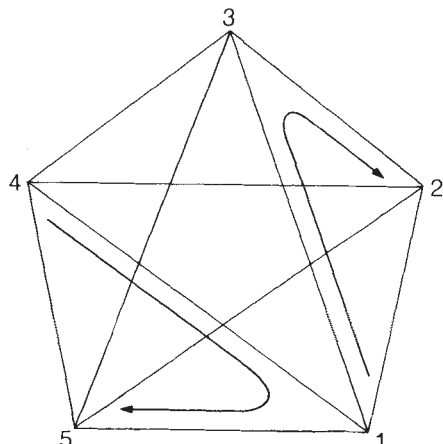
TELECOMMUNICATIONS networks have to choose routes dynamically to make the best use of lines available at each moment. If the most direct link for an attempted conversation or data packet is busy, an exchange tries to find a more roundabout way to get the message through. Complex and ingenious schemes have been proposed for doing this automatically. New work shows that a surprisingly simple scheme is not only very good but comes near to making the best use of transmission capacity that could be achieved by any possible scheme. Frank Kelly and Richard Gibbens, in collaboration with British Telecom researchers, studied the absolute bounds on capacity determined by the nature of the telecommunications network, and it was while thinking about these problems that they arrived at the new method.

To get the bounds they showed that although network traffic is stochastic — calls do not arrive in an exactly predictable way — the capacity problem takes little heed of this. When a major trunk link is nearly saturated it might be carrying 1,000 calls, and the standard deviation of the call distribution will then be about 30, which is relatively small. By solving a deterministic problem of multicommodity flow, Kelly and Gibbens could obtain upper bounds on capacity. This particular problem has special structure, and is not unmanageably large, so it can be solved by linear programming: the difficulties with general multicommodity flow problems (see my recent News and Views article in *Nature* 321, 19; 1986) do not arise here.

The real question is how to get near to these upper bounds. Imagine a trunk network. There will be 50 or so trunk exchanges in the new UK digital network, dealing with long-distance communications traffic in much the same way the national railway services handle long-distance passenger traffic. The trunk network will be fully connected: every exchange is (effectively) connected by its own lines to every other. This means that a call from, say, Manchester to Edinburgh will normally use the direct Manchester-Edinburgh lines. If they are all busy, the Manchester exchange might try to route via Glasgow (or via Southampton, for that matter). If there is no two-link route of this kind, a route with more links might be tried, although there are good reasons why this game quickly becomes self-defeating. An n -link call is a single call using as much capacity as n calls would,

and perhaps causing $n-1$ calls to fail later because there is no route available. As the network became very busy, more and more calls would be routed in this wasteful manner, using a lot of capacity for only a small increase (or even a decrease) in throughput. In other words, n -link routing is unstable.

One-link (direct) routing often gives well below the theoretical network capacity: this is obvious, and is the reason dynamic routing is considered in the first place. Kelly and Gibbens decided to try



A model of dynamic routing. Calls from 1 to 2 overflow via 3; those from 4 to 5 overflow via 1. two-link routing in an attempt to get a significant capacity increase while avoiding instability. The problem was to choose which exchange to act as the middleman in a two-link route when calls are blocked because a direct route is busy. One scheme would be to give every exchange a fixed partner that would be asked to try to forward blocked calls; this solution improves capacity but requires careful traffic analysis and in any case needs frequent reprogramming because traffic patterns change over time. It is also inflexible during, for example, breakdowns and unexpectedly high traffic at the partner exchange.

A very simple way around this problem is the following. An exchange puts calls through directly as long as there are lines available. If all direct lines are busy, it chooses a partner exchange at random and asks it to forward the call. If the chosen partner succeeds, it is chosen the next time a call needs to be forwarded, and keeps getting chosen until it fails. ("All lines from Manchester are engaged. Please try later".) Then a partner is chosen at random again, and so on. It can be seen at

once that this scheme will explore the network until it finds a friendly partner, without any fancy pre-analysis of traffic. A few (very few) calls will be lost, but they would have been lost anyway without the dynamic routing. In the long run, the Manchester exchange will spend most of its time with the friendliest partner. It will adapt itself completely automatically in the face of breakdowns or variations in traffic.

This 'sticky random' selection is a neat trick, but there is another difficulty with this scheme (or indeed with any scheme that allows indirect routing): unselfishness towards one's friends is good up to a point, but sometimes it has to be every exchange for itself! If the Glasgow exchange is as kind about forwarding calls from Manchester as it is about putting through calls from its own subscribers, it will be risking instability even in a system where two-link calls are the worst that are allowed. Somehow, we need to penalize two-link calls, at least when the lines are very busy. One way is to have a trunk-reservation parameter k . If there are at least k free lines on a given link, an exchange forwards calls out along that link without question. If there are fewer than k free lines, it refuses to forward calls at all along that link.

Trunk reservation and sticky random-partner selection turn out to work together in a synergistic way. With sensible values of k , Kelly and Gibbens claim they can get so close to the bound on what is achievable that there is no reason even to bother with other schemes unless they have advantages quite other than improved capacity. Perhaps surprisingly, performance is very insensitive to the exact value of k . As long as k is not equal to zero (total unselfishness, causing instability) or comparable with the channel capacity (complete selfishness, causing inefficiency) the system performs well. For links of a few score up to many thousands of lines, taking k anywhere between 5 and 10 is just fine.

British Telecom, Kelly and Gibbens have applied for a patent for the scheme, under the name Dynamic Alternative Routing. It is possible that the new electronic 'System X' exchanges will be given the capacity to use it, and if it proves as effective in practice as it has been in theory and in simulation, it is likely to become the preferred routing method, for its robust simplicity as much as because of its near optimality. □

Alistair Mees is Professor of Modern Applied Mathematics at the University of Western Australia, Nedlands, Western Australia 6009. The work is awaiting patent and is not yet published. Interested readers are asked to contact Dr Frank Kelly or Mr Richard Gibbens at the Department of Pure Mathematics and Mathematical Statistics, 16 Mill Lane, Cambridge CB2 1SB, UK.

Genetic variability

Clones within a coral reef

from Jared M. Diamond

A LARGE coral reef typically consists of many species, thousands of colonies and millions of polyps. But most corals can reproduce asexually, suggesting that many colonies are genetically identical members of the same clone. How many different genotypes are actually represented among the colonies of one coral species on a reef? A recent attack on this question by Cynthia Hunter^{1,2} opens the way to studying the population genetics of one of the most important groups of marine organisms and will improve our understanding not only of corals but also of many other species of clonal animals.

Exploring the consequences of clonality requires methods for distinguishing members of the same and different clones. Most studies of these questions involve terrestrial plants³⁻⁵; recent histocompatibility studies of graft acceptance or rejection allow identification of putative clones among species of corals⁶⁻⁹ or sponges¹⁰⁻¹². A potential objection to these interpretations is that, whereas almost every individual vertebrate has a unique histocompatibility type, this might not be true of species in which histocompatibility depends on just a few loci. That fear has been realized: electrophoresis sometimes detects differences between individual corals sharing the same histocompatibility type^{9,12-14}. The same objection can be raised against any method of clone identification that samples just a fraction of the genome. Comparing entire genomes is impractical, but the genome should at least be sampled in multiple ways.

Hunter worked with one of the most abundant coral species on Hawaiian reefs, *Porites compressa*, which reproduces sexually once a year by releasing eggs and sperm and may reproduce asexually at any time by colony fragmentation. A host of important conclusions about coral genetic composition, reproduction, age and evolution emerges from her analysis. The area under study included 291 *P. compressa* colonies and millions of polyps, most of which prove to belong to one of only eight genotypes, 40 per cent of them to one of only three genotypes. No colony is smaller than 60 cm², implying that successful larval settlement is rare and that asexual reproduction is far more frequent than sexual reproduction. From measured extents and growth rates of colonies, most colony ages were estimated at 4-21 years, but large colonies (greater than 10 m²) may be hundreds or even thousands of years old^{15,16}. Morphotypes differ in growth rates, competitive ability and distribution between leeward and windward

sides of the reef.

Hunter selected a 2 × 10 m reef area off the Hawaiian island of Oahu and used four methods for identifying putative clones. First, she identified eight morphotypes on the *P. compressa* reef by differences in colour, length and width of branches and by distance between branch tips. That colonies can be visually dis-

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Finger-like colonies of the coral *Allopora nobilis* (courtesy of *Sagittarius*).

tinguished by means of morphological characters is an important advantage of studying *P. compressa*.

Second, she tested histocompatibility by grafting a branch of one colony to a branch of either the same colony, a different colony of the same morphotype, or a colony of a different morphotype. Acceptance or rejection of the graft became evident within a few weeks. All results proved to be transitive: if colony A fused with colonies B and C, colonies B and C also fused with each other. Grafts between colonies that Hunter had assigned to different morphotypes were always rejected, whereas 82 per cent of grafts of colonies assigned to the same morphotypes fused. (The rejected 18 per cent mostly involved two hard-to-distinguish morphotypes classified early in the project, while she was still learning to detect subtle differences between morphotypes.)

Third, Hunter studied seven enzymes by starch-gel electrophoresis. Two proved monomorphic and therefore useless for distinguishing clones, but five were polymorphic. Each morphotype proved electrophoretically distinct, whereas all

colonies of a given morphotype were electrophoretically identical. Most of the 8 × 7/2 = 28 morphotype pairs could be distinguished by means of three or fewer electrophoretic loci, but two of the pairs were identical at four loci and could be distinguished only by the fifth. Hence, for *P. compressa*, unlike some other coral species, electrophoresis and histocompatibility recognize the same sets of clones.

Finally, Hunter used high-performance liquid chromatography to separate seven ultraviolet-absorbing compounds that were methanol-extracted from coral tissues. The eight morphotypes could all be distinguished by their characteristic ultraviolet signatures, and colonies of the same morphotype proved identical.

Thus, each morphotype is unique in its histocompatibility traits, 5-locus genotype and ultraviolet signature. Within Hunter's 20 m² transect, one morphotype was represented by only a single colony, but the other morphotypes were multiply represented (by up to 38, and an average of 19, colonies). During the two years of Hunter's study the dominant colonies increased in area, the rare ones retreated. Genetic variability may itself exhibit great local variation: it may be maximal on new substrates available for colonization by sexually produced propagules, and minimal on old stable reefs where colonies are established mainly by fragmentation⁸.

To appreciate the significance of Hunter's work, reflect that virtually every individual of a species practicing exclusively sexual reproduction (such as most vertebrates) represents a unique genotype. But many plants and invertebrates, plus a few vertebrates, reproduce asexually as well as sexually. The fitness of a genotype then depends on all the asexually produced ramets constituting the clone. As summarized recently by Jackson³, "clonal growth — the formation of more than one individual of identical genetic composition — is a fundamental ecological adaptation that has far-reaching consequences for the population biology, morphology, development and evolution of such organisms". □

1. Hunter, C.L. *Proc. Fifth Int. Coral Reef Congr.* 6, 69 (1985).
2. Hunter, C.L. *Tester Symp. Abstr.*, Honolulu 22 (1986).
3. Jackson, J.B.C. *et al. Population Biology and Evolution of Clonal Organisms* (Yale University Press, New Haven, 1985).
4. Aarssen, L.W. & Turkington, R. *J. Ecol.* 73, 605 (1985).
5. Hartnett, D.C. & Bazzaz, F.A. *J. Ecol.* 73, 407 (1985).
6. Hildemann, W. *et al. Immunogenetics* 2, 269 (1975).
7. Jokiel, P. *et al. Bull. Mar. Sci.* 33, 181 (1983).
8. Neigel, J.E. & Avise, J.C. *Evolution* 37, 437 (1983).
9. Willis, B. & Ayre, D.J. *Oecologia* 65, 516 (1985).
10. Van de Vyver, G. *Archs zool. exp. Gen.* 112, 55 (1970).
11. Kaye, H. & Ortiz, T. *Mar. Biol.* 63, 165 (1981).
12. Curtis, A.S.G. *et al. Transplantation* 33, 127 (1982).
13. Stoddart, J.A. *et al. Evolution* 39, 461 (1985).
14. Heyward, A.J. & Stoddart, J.A. *Mar. Biol.* 85, 117 (1985).
15. Nozaki, Y. *et al. Geophys. Res. Lett.* 5, 825 (1978).
16. Druffel, E.M. *Science* 218, 13 (1982).

Jared M. Diamond is Professor of Physiology at the University of California Medical School, Los Angeles, California 90024, USA.

Molecular biology

A new protein in petunia

from J.E. Varner and G.I. Cassab

CAROL Condit and Rich Meagher, on page 178 of this issue¹, report the discovery of a gene in petunia plants that is predicted to encode a protein containing 67% glycine residues with a sequence mainly conforming to the formula (glycine-X)_n, where X is frequently glycine. The product of the gene is presumably secreted because it begins with a transit peptide sequence. Condit and Meagher propose that the protein is secreted into the cell-wall compartment where it forms part of the extracellular matrix.

Speak of glycine-rich proteins and one calls to mind structural proteins made up of short repeating units, for example collagen (X-proline-glycine)_n, the cocoon silks *Bombyx mori* (X-glycine)_n and *Antheraea paphia* (25% glycine); and spider-web silks (41–52% glycine)². In collagen the presence of glycine at every third position in the peptide chain, together with the presence of proline or hydroxyproline at every second position, allows three peptide chains to form the triple-helical structure that gives collagen its unique and useful properties. The occurrence of glycine at every fourth or second residue in silks allows the formation of parallel and antiparallel β -sheets. These glycine-rich conformations meet the structural and architectural needs of many species.

As it happens there are several known instances in plant tissues in which glycine is a major fraction of the total protein nitrogen. These include the soybean (*Glycine max*) seed coat (21% glycine)³, the gourd (*Cucurbita ficifolia*) seed coat (21% glycine)⁴, the pumpkin (*Curcubita pepo*) seed coat (the major protein of the cell walls contains more than 47% glycine; our own unpublished observations), milkweed (*Periploca graeca*) stem (cell walls, 31% glycine)⁵, moisture-stressed soybean hypocotyls (cell walls, the major soluble protein contains 25% glycine; C.S. Bozarth, J.E. Mullett and J.S. Boyer, personal communication) and oat (*Avena*

sativa) coleoptiles (epidermal cell walls, 27% glycine; D. Pope, personal communication). Because cell walls contain several enzymes with normal content of glycine (8–12%) and a hydroxyproline-rich structural glycoprotein that contains no glycine, we are certain that for the tissues mentioned the glycine-rich proteins, when purified and characterized, will be found to be two to three times as rich in glycine as the cell walls they are extracted from. To cite another example of the importance of glycine richness in structural proteins, one of the glycoproteins which makes up *Clamydomonas reinhardtii* cell walls is glycine-rich (23%)⁶. And, finally, to illustrate the usefulness of glycine-rich proteins in cell-wall architecture, *Thermomicrobium roseum*, a Gram-

negative, obligately thermophilic bacterium, possesses a cell wall that is composed predominantly of a protein with a monomeric relative molecular mass of 75,000 that is 34% glycine⁷.

In the past few years there has been a general and belated realization that the hydroxyproline-rich glycoproteins of plant cell walls discovered by Lamport and Northcote⁸ in 1960 are of considerable importance in cell-wall structure (particularly in dicotyledons), not only with respect to growth and development⁹ but also in the response of the plant to injury and infection¹⁰. The discovery and characterization of a gene encoding a protein remarkably rich in glycine that is probably localized in the wall is likely to increase further our awareness of cell-wall proteins. Perhaps we should now ask whether the plant cell wall contains other kinds of structural proteins. □

J.E. Varner and G.I. Cassab are in the Department of Biology, Washington University, St Louis, Missouri 63130, USA.

Immunology

Bone marrow grafts and tolerance

from E. Donnall Thomas

BONE marrow transplantation is now widely used in human beings as a treatment for both acquired and genetically determined diseases. However, transplantation between individuals who are not genetically identical may give rise to an immunological reaction of host cells against donor cells resulting in graft rejection, or of graft cells against the host resulting in graft-versus-host disease. T lymphocytes, or T-cell subsets, are thought to mediate these reactions. Graft rejection is uncommon in recipients receiving intensive chemoradiotherapy before grafting but the problems of graft-versus-host disease remain and may be severe even if donor and recipient are matched for the antigens of the major histocompatibility system, HLA. In this issue (*Nature* 323, 164; 1986) Cobbold *et al.* describe the use of monoclonal antibodies against T cells in mice that may point the way to the resolution of these problems.

Immunosuppressive drugs or drug combinations have been developed to prevent graft-versus-host disease in most patients with matched or partially matched donors (Storb *et al. N. Engl. J. Med.* 314, 729; 1986) but they are not always effective and are associated with significant toxicity. For this reason, investigators have studied various methods of removing T cells from the donor marrow before the marrow is given to the patient. These approaches have resulted in a decreased incidence and severity of graft-versus-host disease, but a

new problem has arisen, the problem of graft failure within a few weeks or months after the transplant.

The reasons for graft failure are not understood but include the possibility of damage to haematopoietic stem cells during the T-cell depletion; the possibility that T cells are in some way essential to normal haematopoiesis; and the possibility that the absence of donor T cells permits an immunological reaction of residual host T cells resulting in graft rejection. The last possibility is supported by the demonstration of small numbers of residual host T cells after treatment by the use of cytogenetics and analysis of restriction fragment length polymorphisms.

The study by Cobbold *et al.* in a murine model of bone marrow transplantation across the major histocompatibility barrier shows clearly that an immunological reaction by T cells is the mechanism of graft rejection and that it involves more than one subset of T cells. The authors administered monoclonal antibodies reacting with both the L3/T4⁺ and Lyt-2⁺ T-cell subsets (equivalent to the CD4⁺ and CD8⁺ subsets in man) that prevented graft rejection even after sublethal irradiation exposure. Surprisingly, and of greatest interest, many of these animals became long-lived chimaeras, specifically accepting a donor's skin graft, that is, they were operationally tolerant.

A preliminary study by Fisher *et al.* (*Bone Marrow Transplant* 1, 167; 1986)

1. Condit, C. & Meagher, R. *Nature* 323, 178 (1986).
2. Gosline, J.M., DeMont, M.E. & Denny, M.W. *Endeavour* 10, 37 (1986).
3. Rackis, J.J., Anderson, R.L., Sasame, H.A., Smith, A.K. & Van Etten, C.H. *Agric. Food Chem.* 9, 409 (1961).
4. Dreher, M.L., Weber, W., Bemis, W.P. & Berry, J.W. *J. Agric. Food Chem.* 28, 364 (1980).
5. Melin, D., Vallee, J.C., Vansuyt, G. & Prevost, J. *Physiol. Plant.* 46, 68 (1979).
6. Goodenough, U.W., Gebhart, B., Mecham, R.P. & Heuser, J.E. *J. Cell Biol.* (in the press).
7. Merkel, G.J., Durham, D.R. & Perry, J.J. *Can. J. Microbiol.* 26, 556 (1980).
8. Lamport, D.T.A. & Northcote, D.H. *Nature* 188, 665 (1960).
9. Cassab, G.I., Nieto-Sotelo, J., Cooper, J.B., Van Holst, G.J. & Varner, J.E. *Plant Physiol.* 77, 532 (1985).
10. Showalter, A.M. & Varner, J.E. in *Molecular Strategies for Crop Protection* (eds Arntzen, C. & Ryan, C.) (Liss, New York, in the press).

describes the prevention of graft failure in patients with immunological deficiency diseases given HLA-incompatible bone marrow transplants depleted of donor T cells. They used an anti-LFA1 (CDw18 in humans) monoclonal antibody that prevents cell adhesion and therefore interferes with the immunological reactions of a broad spectrum of immunologically competent cells.

These observations in mouse and man point the way to a resolution not only of the graft rejection problem but also resolution of the graft-versus-host disease problem. However, we must keep in mind that we cannot alter the normal human donor by such procedures as thymectomy and antibody administration as was done in the murine model of Cobbold *et al.*

Will murine monoclonal antibodies have sufficient distribution and activity

in human patients or will monoclonal antibodies of human or other primate origin be necessary? Will depletion of donor T cells or administration of anti-lymphocyte monoclonal antibodies *in vivo* diminish immunological competence and result in a greater likelihood of recurrence of disease in leukaemia patients because of the lack of an anti-leukaemia immune response? Will resistance against viral illnesses fail to develop? Many biological functions require more than tolerance — they require cooperation between different cells and tissues and may be perturbed in ways not yet anticipated by these genetic disparities and manipulative procedures. □

E. Donnell Thomas is in the Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, Washington 98104, USA.

Earth sciences

Is there water in the deep continental crust?

from Bruce W.D. Yardley

RECENT developments in the earth sciences provide remarkable insights into the composition of the Earth's interior and of other bodies in the Solar System, but uncertainty remains about the make-up of other parts of the Earth. The question of whether or not the continental crust, especially its deeper parts, contains water as a discrete phase is important for understanding geological processes, but is controversial because of contradictory inferences from the physical and chemical lines of evidence. D.I. Gough, on page 143 of this issue¹, puts forward a new model that seeks to meet the geophysical constraints. Is the model consistent with the chemical data?

Conventional wisdom has it that, away from areas actively undergoing metamorphic heating, the lower crust (below 15–20 km) is essentially dry with no free fluid phase present, although small amounts of hydrous minerals or carbonates may persist. This view is based on the mineralogy of rocks derived from the lower crust and now exposed in high-grade metamorphic terranes or found as volcanic xenoliths².

The argument for a dry lower crust is as follows: the granulite facies assemblages that predominate in lower crustal rocks are stable in the presence of an aqueous phase, if at all, only at the extreme metamorphic temperatures at which the rocks originally formed, whereas the present-day temperatures of the lower crust are generally less than this. If water penetrates granulites at normal crustal temperatures it will be entirely consumed by hydration reactions forming lower-grade,

hydrous minerals³, until the granulite facies assemblage is entirely retrograded. Furthermore it is clear both from experimental studies⁴ and from reactions in geothermal fields that the hydration reactions are fast, so that the possibility of unstable persistence of water in the lower crust can be excluded. Hence the occurrence of granulite facies assemblages suggests that the lower crust remains dry and any water gaining access becomes combined into hydrous minerals.

This model does not account for the geophysical differences between the upper and lower crust, and is apparently at variance with the occurrence of anomalously high electrical conductivities at depths of 15–20 km that have been widely documented in Europe and North America and have been ascribed by several workers to the presence of an aqueous fluid phase. Gough's model¹ explicitly accounts for the major geophysical characteristics of the upper and lower crust on the basis of the different modes of occurrence of aqueous fluid in the two regions, rather than the lithological differences assumed in studies based on rock samples⁵. Gough points out that the lower crust is anomalously conductive, contains seismic reflectors and, at least in tectonically active areas, deforms in a ductile manner, whereas the crystalline upper crust is resistive, brittle and has few seismic reflectors. He suggests that this contrast can be explained if water is present as a continuous film in the lower crust but at shallower depths and lower temperature is isolated in discrete cavities held closed in

the normal upper crustal stress fluid.

Does Gough's model satisfy other lines of evidence about the nature of the crust, or will it fail in the same way that the 'dry granulite' model appears to fail to meet the geophysical constraints? First, the different approaches to the lower crust do not necessarily measure quite the same thing. Geophysical measurements refer to the lower crust as it is today, whereas lower crustal rock samples have been brought to the surface often quite early in their geological history. Second, much of the information obtained from rock samples refers to their original formation rather than to the time they may have spent as stable lower crust. Despite this caveat it still appears that the geochemical arguments for a dry lower crust do preclude the existence of a free aqueous phase, unless it is restricted to totally hydrated layers within a predominantly dry crust, or unless lower crustal temperatures commonly approach those of granulite facies metamorphism. This latter suggestion does not seem likely when the metamorphic histories of granulite terranes are considered.

In the upper crust, there is in fact direct evidence for fluids flowing, most spectacularly from the deep borehole on the Kola peninsula⁶, where influx of water is recorded to at least 12 km. It may also be significant that the most conspicuous features of road-cuts and quarries in crystalline rocks are often joints lined with hydrous minerals.

Clearly, there must be a flaw in one or both of the arguments if geophysically based models are so at variance with those derived from geochemical arguments. Both cases depend on an adequate understanding of the physical or chemical properties of rocks at high pressures and temperatures, where experimental data are necessarily restricted. Although there are experimental measurements of the seismic velocities of xenolith suites that confirm their lower crustal provenance⁵, there are little data on electrical conductivity at deep crustal conditions. Tantalizingly, what little data that there are⁷ suggests that basic granulites have a higher conductivity at elevated pressure and temperature than can be accounted for by the pore fluid that was present in the experiments alone. If this is confirmed by further work, the geophysical and geochemical viewpoints could be reconciled with a 'dry' lower crustal model. □

1. Gough, D.I. *Nature* **323**, 143 (1986).

2. Kay, R.W. & Kay, S.M. *Rev. Geophys. Space Phys.* **19**, 271 (1981).

3. Yardley, B.W.D. *Geology* **9**, 405 (1981).

4. Wood, B.J. & Walther, J.V. *Science* **222**, 413 (1983).

5. Jackson, I. & Arculus, R.J. *Tectonophysics* **101**, 185 (1984).

6. Kozlovsky, Ye. A. *Scient. Am.* **251**(12), 106 (1984).

7. Lee, C.D., Vine, F.J. & Ross, R.G. *Geophys. J. R. astr. Soc.* **72**, 353 (1983).

Bruce W.D. Yardley is in the Department of Earth Sciences at the University of Leeds, Leeds LS2 9JT, UK.

Archaeology

A Tuscan mining village

from Richard Hodges

RECONSTRUCTIONS of mediaeval village life conjure up an image of harsh, feudal conditions for an undernourished peasant community. The remarkable contemporary descriptions of the fourteenth century Pyrenean village of Montailou suggest badly constructed dwellings, a poor diet and long hard winters (see Ladurie, E. Le R. *Montailou* Gallimard, Paris, 1975). Therefore, it comes as a surprise to discover that Mediterranean coastal villages of the same date are far from primitive. Excavations at the deserted mediaeval hill-top village of Rougiers near Aix-en-Provence (D'Archimbaud, G.D. *Les Fouilles de Rougiers* CNRS, Paris, 1980), revealing a great array of imported ceramic tablewares, glassware and fine metalwork, convincingly counter the impression given by Ladurie. But is Rougiers an exception? Riccardo Francovich's current large-scale excavations of the mining village of Rocca San Silvestro, near Piombino on the coast of Tuscany, Italy, not only endorse the discoveries at Rougiers but offer us a mediaeval archaeological reconstruction to match the description of Montailou (Francovich, R. *et al. Arch. Medievale* 12, 313; 1985).

Rocca San Silvestro lies in a mining zone that has origins in the Etruscan period, but today is suffering from industrial malaise. The first reference to it occurs in a document of 1108, and thereafter it figures commonly in twelfth and thirteenth century Pisan sources. Several cartularies record it as a mining village (40–45 houses supporting 230–260 people) associated with iron and copper. But by the early fifteenth century it had been long abandoned, falling victim to the great fourteenth century economic depression in Tuscany and losing out to villages nearer the sea.

In 1984, clearance of the vegetation concealing much of the hilltop revealed the well-preserved dwellings clustered beneath a small tower (see figure). Excavations in 1985 showed that the buildings are remarkable not just for their preservation, but for the quality of the construction. Most of them were probably built in the twelfth century and were never repaired. This is most unusual: at Rougiers, for example, most of the deserted structures are ostensibly late medieval rebuildings of earlier dwellings. The houses and the church have well-coursed stonework while, together with the cobbled streets, worn door-thresholds and paved floors would not be out of place in present-day villages in the region.

The prospect of documenting early

Renaissance mining techniques as well as the mines close to the village constitute an important part of Francovich's project. He found a thirteenth century iron adit mine in a building just outside the walls of the settlement, and located copper smelting in an unoccupied part in the northern part of the settlement. The excavations of these areas will continue, together with the bid to pinpoint the precise sources of the metal ores by spectroscopic techniques. But, as Francovich notes, it should be possible to quantify the output of metals by the village and to simulate its economic history to shed new light on the bare register of its written history.

Rocca San Silvestro is hardly known, but these excavations and the analysis of its industries should gain it a place in mediaeval history alongside Montailou. The discoveries reveal the early investment made in mining copper and iron ores



Riccardo Francovich

—iron for ships' nails, tools and weapons; and copper for the myriad kinds of jewellery that were the fashion in hill villages and in the great seats of the Renaissance such as Florence and Siena. The minerals of this region were important in the rise of the Etruscan civilization and for the city communes of the late Middle Ages. But, like mining villages throughout the history of western Europe, the zenith of Rocca San Silvestro was short lived. □

Richard Hodges is in the Department of Archaeology and Prehistory at the University of Sheffield, Sheffield S10 2TN, UK.

Biochemistry

G-protein control of inositol phosphate hydrolysis

from Bob Michell and Chris Kirk

THE many cell-surface receptors that regulate adenylate cyclase do so through two guanine nucleotide-dependent G proteins (or N-proteins), one of them stimulatory (G_s) and the other inhibitory (G_i). Recent work has revealed that these are two members of a large family of G proteins that are similar in structure and are essential to a variety of signal-transducing reactions at membranes (see the recent discussion in *News and Views*¹). A second major family of receptors transmits its message into target cells by activating a phosphoinositidase C (PIC) that hydrolyses phosphatidylinositol 4,5-bisphosphate (PtdIns(4,5) P_2) to 1,2-diacylglycerol and inositol 1,4,5-trisphosphate (Ins(1,4,5) P_3). These compounds both act as intracellular messengers (reviewed in refs 3–5). It is now clear that GTP is essential for PIC activation^{6–8}, but the putative G protein involved in this process has not been identified. A report from Wakelam and co-workers on page 173 of this issue⁹ provides the strongest of several recent indications that the GTP-binding p21 proteins encoded by *ras* (proto)oncogenes may be involved in this process.

The preliminary identification of G proteins responsible for receptor-effector coupling in new transmembrane signalling

systems frequently uses bacterial toxins that attach ADP-ribose to the α -subunits of particular G proteins and modify their activities. However, attempts to identify a G protein essential to PIC activation by this method have been confusing rather than helpful. In some cells, receptor-PIC coupling is inhibited by pertussis toxin (islet-activating protein), suggesting a role for G_i (a G protein of unassigned function) or G_o (for example, refs 10–12). But pertussis toxin is without effect on PIC activation in other cells, even though it inhibits the action of G_i on adenylate cyclase in these cells (for example, refs 13, 14). Similarly, cholera toxin, which has been used to identify G_s -mediated reactions, has different effects on receptor-PIC coupling in different cells (refs 14, 15; unpublished data of G. Guillon *et al.*). Thus it seems clear that none of the well-characterized G proteins can be universally responsible for mediating receptor-PIC coupling. An obvious but experimentally daunting possibility is that different but closely related G proteins may mediate this process in different cells and/or at different receptors.

The idea that the p21 G proteins encoded by the three closely related *ras* proto-oncogenes may mediate recep-

tor-PIC coupling was first raised when it was realized that PIC activation is important in control of growth by normal proliferative stimuli and by oncogenes^{4,5,9}. The first evidence was obtained by Chiarugi *et al.*¹⁶, who showed that transformation of 3T3 fibroblasts by a transforming *ras* gene from a human bladder cancer increases the responsiveness of PIC to muscarinic cholinergic stimulation. Fleischman *et al.*¹⁷ then showed that fibroblasts stably transformed by the three types of *ras* oncogenes (*Ha-ras*-1, *Ki-ras*-2 and *N-ras*) always display a steady-state increase in the concentrations of 1,2-diacylglycerol and inositol phosphates relative to their precursor (PtdIns(4,5)P₂), indicating a sustained PIC-mediated turnover of PtdIns(4,5)P₂. A similar sustained increase in 1,2-diacylglycerol was seen by Preiss *et al.*¹⁸ in cells infected with a temperature-sensitive transforming *Ki-ras* gene and grown for four days at permissive temperature. In both of these studies, cells were maintained in media that included the growth factors of fetal serum, so continuous PIC activity might have been a consequence either of constitutive activation of PIC by transforming *ras*-encoded p21 G proteins or of an amplified response of PIC to stimulation of growth factor receptors.

Wakelam *et al.*⁹ now show that an amplification of the PIC response to growth factors can also be produced by a controlled overexpression of p21^{N-ras}, the protein that is encoded by the *N-ras* proto-oncogene of normal, non-transformed cells. This protein only causes phenotypic transformation when grossly overexpressed. The authors used a gene construct that allows regulation of the cellular concentration of p21^{N-ras} by dexamethasone, and showed that its over-expression leads to parallel increases in the growth-promoting and PIC-activating responses to bombesin, bradykinin, vasopressin, fetal serum and platelet-derived growth factor. They therefore suggest that p21^{N-ras}

may be (one of) the coupling protein(s) responsible for receptor-PIC coupling, and that in normal cells the concentration of this protein is an important factor that regulates the intensity of signalling through the PIC pathway.

Although this is the simplest interpretation of the available information, it must be regarded as tentative for the present. An unexplained feature of the new data⁹ is that over-expression of p21^{N-ras} also potentiates growth stimulation by epidermal growth factor (EGF), a growth factor that does not activate PIC in fibroblasts: this might be a direct effect of p21^{N-ras} on EGF signalling or the result of synergism with a PIC response to other agents. In all the

experiments mentioned above, *ras* transformation or over-expression has been maintained for two days or more, and it is quite possible that this causes the accumulation of other gene product(s) essential for control of PIC. This uncertainty might, in principle, be resolved by using temperature-sensitive *ras* mutants that allow quicker changes in the p21 status of cells. Unfortunately, the only published experiments of this type¹⁸ were undertaken over much too long a timescale to clarify this issue. □

Bob Michell and Chris Kirk are in the Department of Biochemistry at the University of Birmingham, Birmingham B15 2TT, UK.

Georgyi Frantsevich Gause 1910–1986

GEORGYI FRANTSEVICH GAUSE, who died on 2 May 1986, belonged to two worlds in biology. His early ecological and evolutionary papers were almost unknown in the antibiotic community, and ecologists thought that he was lost to biology when he turned to antibiotics. In fact he made outstanding contributions in both areas.

Gause is probably best known for his classical monograph *The Struggle for Existence*, published in 1934, which summarized experiments on the competition between different species of protists in an experimentally defined system, enabling him to formulate the principle of competitive exclusion. According to this principle two species belonging to one ecological niche cannot coexist for any length of time.

In 1927 Gause became a student of biology at Moscow University and his mentor and friend Professor V. V. Alpatov took up a two-year Rockefeller fellowship in Baltimore to work with Raymond Pearl. While Alpatov was absent Gause was influenced by Professor Evgeny Smirnov, a controversial larmarkian and an advocate of mathematical methods in taxonomy. Gause's studies on the distribution and ecological plasticity of Orthoptera in North Caucasus were among the first using statistical methods in field biological research.

The return of Alpatov from Baltimore brought the spirit of experimentalism to the Moscow school of biologists, creating the intellectual background for Gause's studies which he described in *The Struggle for Existence*. Perhaps his most interesting experiments were those on protist species belonging to one or two niches. Gause had already experimentally demonstrated conditions leading to coexistence or exclusion of the competing species, and now turned his attention to the predator-prey system. His experiments in test-tubes (which Gause called microcosms) were relevant to processes occurring in the biosphere.

Gause contributed to further developments in the mathematical theory of relations between species. In 1935 he published a paper with A. A. Witt (a well

known mathematician and physicist) in *American Naturalist* which extended mathematical models of relations between the species to include commensalism, mutualism and symbiosis. The full significance of this paper was not realized until the 1970s.

At about the same time Gause became interested in the action of D- and L- isomers of various substances on biological systems, and he summarized his studies on dissymmetry of protoplasm and selective toxicity of D- and L- isomers in his book *Optical Activity and Living Matter* in 1941. It was probably his interest in biological mechanisms associated with the struggle for existence in the world of microbes that led Gause to study of antibiotics.

In 1942 Gause and his wife Maria Georgyevna Brazhnikova isolated a strain of *Bacillus brevis* that produced a bacterial antibiotic substance. This was later named Gramicidin S and was soon used at military hospitals to treat infected wounds and became popular in various structural and functional studies. For this work Gause received the Stalin prize in 1946. The laboratory of antibiotics founded by Gause and Brazhnikova in 1942 was reorganized into an institute and from 1960 until his death Gause was its director.

In the search for antibiotics he followed the ecological approach of his younger years. He and his associates developed principles for the classification of actinomycetes which became widely recognized in the search for new biologically active substances. Gause's results were impressive. His institute developed many antibacterial and antitumour antibiotics used in the treatment of infections and malignancies.

From one of the founders of mathematical ecology Gause evolved into a world figure in antibiotic research. His works are an example what can be achieved by simple technical means, clear thinking, imaginative planning and the right choice of system.

Nikolai N. Vorontsov and Jakov M. Gall

1. Gilman, A.G. *Cell* **36**, 577 (1984).
2. Bourne, H.R. *Nature News and Views* **321**, 814 (1986).
3. Nishizuka, Y. *Nature* **308**, 693 (1984).
4. Berridge, M.J. & Irvine, R.F. *Nature* **312**, 315 (1984).
5. Downes, C.P. & Michell, R.H. In *Molecular Mechanisms of Transmembrane Signalling* (eds Cohen, P. & Houslay, M.D.) **3** (Elsevier, Amsterdam, 1985).
6. Haslam, R.J. & Davidson, M.M.L. *J. Receptor Res.* **4**, 605 (1984).
7. Cockcroft, S. & Gomperts, B.D. *Nature* **314**, 534 (1985).
8. Litoch, I., Wallis, C. & Fain, J.N. *J. biol. Chem.* **260**, 5464 (1985).
9. Wakelam, M.J.O. *et al.* *Nature* **323**, 173 (1986).
10. Nakamura, T. & Ui, M. *J. biol. Chem.* **260**, 3584 (1985).
11. Murayama, T. & Ui, M. *J. biol. Chem.* **260**, 7226 (1985).
12. Higashida, H., Streety, R.A., Klee, W. & Nirenberg, M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 942 (1986).
13. Martin, M.W., Evans, T. & Harden, T.K. *Biochem. J.* **229**, 539 (1985).
14. Mettritt, J.E., Taylor, C.W., Rubin, R.P. & Putney, J.W. Jr *Biochem. J.* **236**, 337 (1986).
15. Brandt, S.J., Dougherty, R.W., Lapetina, E.G. & Nield, J.E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3277 (1985).
16. Chiarugi, V., Porciatti, F., Pasquali, F. & Bruni, P. *Biochem. biophys. Res. Commun.* **132**, 900 (1985).
17. Fleischman, L.F., Chahwala, S.B. & Cantley, L. *Science* **231**, 407 (1986).
18. Preiss, J. *et al.* *J. biol. Chem.* **261**, 8597 (1986).

Geometry

Non-euclidean kaleidoscopes

from Ian Stewart

A FRIEND of mine has a brass kaleidoscope. At one end is a rotating disk that produces patterns of breathtaking clarity and beauty using only two angled mirrors — vivid proof that simplicity and beauty go hand in hand. There is an equally beautiful mathematical theory of kaleidoscopes in any number of dimensions, called Coxeter groups (a set of transformations is a group if the result of performing any two in succession remains in the set). In a euclidean space there are non-trivial kaleidoscopes in any number of dimensions, but no such result is known in the non-euclidean case. Now E.B. Vinberg has discovered why (*Trans. Moscow Math. Soc.* **47**, 75; 1985), proving that in high-dimensional spaces with Lobachevskii (non-euclidean) geometry, kaleidoscopes do not exist.

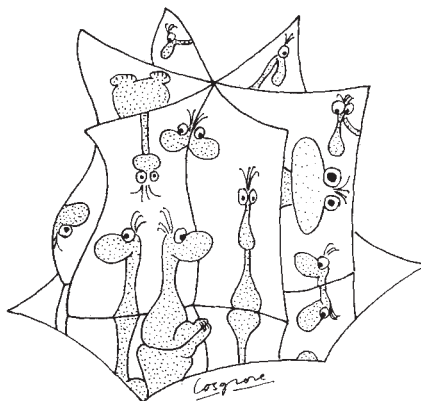
Consider a plane, in which is drawn a straight line acting as a mirror. To each point on the plane corresponds another, its reflection. The operation of reflection is a rigid motion of the plane. Now consider a primitive kaleidoscope: two such mirrors set at an angle. To each corresponds a reflectional motion of the plane; and these motions can be combined by reflecting in either of the mirrors several times. The kaleidoscope effect occurs when a reasonably large but finite (or at least discrete) set of points can be so obtained from any initial point.

When is the number finite? Some integer multiple of the angle between the two mirrors must amount to an exact number of full turns, otherwise the images of a point will be densely distributed around a circular arc. Thus, the angle must be $2\pi p/q$ radians, where p and q are integers, and the requirement of finiteness on the number of images implies arithmetical conditions on the mirror geometry.

The corresponding classification for kaleidoscopes formed by an arbitrary system of mirrors in n -dimensional euclidean space was first obtained by H.S.M. Coxeter in 1934, and it can be found in chapter XI of his book *Regular Polytopes* (2nd edn; Macmillan, London, 1936). He considered not just finite groups, but discrete groups (where the images of a point do not cluster together), and classified them as combinations of a system of basic types. The finite groups fall into four infinite families, together with six exceptions. One family is the groups that use the n -coordinate planes of n -space mirrors, and the others are similar. The exceptions are symmetry groups of unusual regular solids: the dodecahedron in 3-space, the 600-cell in 4-space, and so on. The infinite

discrete groups come as four families plus six exceptions. Each should be thought of as a fundamental geometrical object, likely to show up occasionally in a totally different context simply because there is nothing else quite like it.

Coxeter obtained his classification by associating a graph (a set of nodes joined by edges) to each system of mirrors. The nodes correspond to the mirrors; the number of edges between two nodes correspond to the angles between the associated mirrors. Coxeter obtained restrictions on the possible form of such graphs,



which allowed him to use a combinatorial argument to find all the possibilities.

So much for euclidean space. What about non-euclidean? The idea of non-euclidean geometry is to find self-consistent geometries in which Euclid's parallel axiom fails, but other axioms remain true. (The parallel axiom states that given a straight line and a point not lying on it, there exists a unique parallel through the point.) Such geometries are best imagined by constructing models within euclidean geometry. For example, hyperbolic geometry, generally thought to be the most interesting, can be thought of as follows. Take a circular disk in the euclidean plane and imagine that only the interior of this disk exists. Define a straight line in the disk to be an arc of a circle that meets the boundary at right angles. Then all of Euclid's axioms hold, except that there are many parallels to a given line — just imagine circles that do not intersect.

Henri Poincaré gave a fanciful description of such a non-euclidean world, imagining it to be inhabited by creatures that shrink in size because of a temperature change as they move towards the boundary. He pointed out that it will take them infinitely many steps to reach the boundary, which to them therefore does not exist. A similar picture holds for n -dimensional hyperbolic geometry, except that

now the disk is replaced by an n -dimensional ball. As Poincaré also pointed out, if we ourselves lived in such a world we might have a hard time detecting the shrinkage, especially if it were very slight.

In hyperbolic geometry there are reflections too, so one can ask whether any interesting kaleidoscope groups exist. In the course of work on transformation groups in complex function theory, Poincaré and Felix Klein solved this problem in the two-dimensional case (where there are many possibilities). E.M. Andreev (*Mat. Sbornik* **81**, 445 and **83**, 256; 1970) obtained the classification in the three-dimensional case.

To state Vinberg's result accurately we need one more idea. Consider a tiled floor: select one tile, and place a mirror along each of its sides. Then, under repeated reflections, it will produce every other tile on the floor. This is an infinite kaleidoscope group, and the tile that generates it is called a fundamental domain. This game can be played in hyperbolic geometry. A kaleidoscope group is defined as crystallographic if it has a fundamental domain of finite volume.

Vinberg's main result is that in hyperbolic geometry of dimension 30 or greater there are no crystallographic kaleidoscope groups. The proof once more proceeds by reducing geometric statements to conditions on graphs and amounts to an elaborate generalization of Coxeter's original methods. If the dimension is raised to 62 the proof can be made fairly simple; otherwise the combinatorial complications make life much harder.

In mathematics, an unusual gadget can often be very useful. One way to find new gadgets is to try to produce an exhaustive list of all objects of a given kind, and see whether anything interesting turns up. In this sense the importance of Vinberg's theorem is that it is about the non-existence of gadgets. Such a theorem is far more useful than it looks. To give a imaginary example, suppose yourself to be working on the notorious McGonagall Conjecture, that every anachimedean hypercycle is left-divertible. (There is — I hope — no such conjecture.) It is common to try to prove a theorem by assuming it to be false and deducing a contradiction. You therefore assume that there exists an anachimedean hypercycle that is not left-divertible. After 200 pages of close analysis you find that this implies the existence of a kaleidoscope group in 1,086-dimensional space. (Less likely things have occurred.) Vinberg's result says this cannot happen, so the falsity of the theorem is impossible. That is, the McGonagall Conjecture is true. Knowing what cannot happen is just as important as knowing what can. □

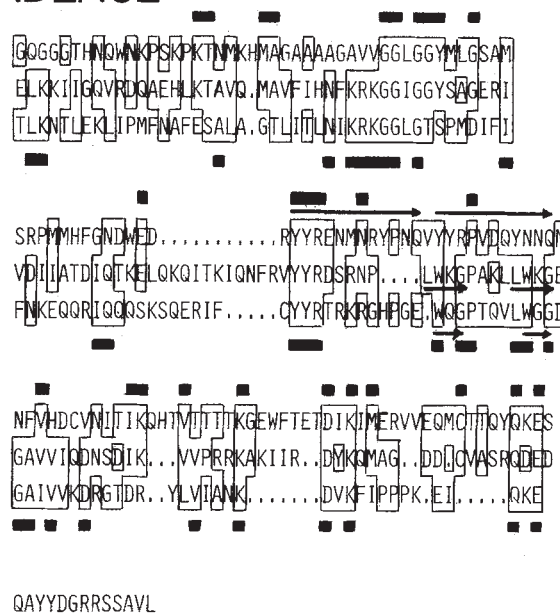
Ian Stewart is in the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

AIDS virus and scrapie agent share protein

SIR—Infection with the human T-lymphotropic virus type III (HTLV-III/LAV or HIV) can result in central and peripheral nervous system disorders in addition to severe immunological defects (see ref. 1 for review). The slow onset and progressive nature of HTLV-III associated neurological disorders is reminiscent of several other transmissible central nervous system disorders, notably kuru, Creutzfeldt-Jacob disease and Gerstman-Straussler in man, and scrapie in sheep. A protein, PrP 27-30, is associated with scrapie agent infectivity² and is related to the product of a mammalian gene, cDNAs from which have been sequenced^{3,4}. The possibility that the neurological symptoms associated with HTLV-III infection are induced by a scrapie-like agent and the previously documented observation that amyloid-like fibrils, similar to those that accompany scrapie, are present in the brain tissues of some AIDS (acquired immune deficiency syndrome) patients⁵, prompted a comparison, reported here, of the sequence of the HTLV-III provirus with that of the hamster gene for PrP 27-30.

Using dynamic biosequence comparative methods (program LOCAL⁶) four subsequences within a 400-nucleotide span of the 3'-end of the polymerase open reading frame (*pol*) of HTLV-III⁷ were found to have a primary structure that is similar to that of the hamster gene for PrP 27-30 (ref. 4) as shown in Fig. 1a. The four subsequences are colinear (Fig. 1b) except for a difference in spacing, part of which can be accounted for by the presence of a triple direct repeat (labelled R3 in Fig. 1b). The subsequences span the entire coding region of the mature PrP 27-30 related protein^{2,4} and can be identified in

Fig. 2 Alignment of the amino acid sequences of the scrapie-related PrP 27-30 protein and the carboxyl-end of the *pol* polyproteins of HTLV-III/LAV (884-1,015) and visna (983-1,105) viruses. Common amino acids and conservative substitutions are boxed and positions with identical amino acids are shown with black bars. There are three subsequences with primary sequence similarity at the end of the first (20% identical amino acids, 25 positions total) and second (20 and 40% identical amino acids, 29 positions total) lines and throughout the third (34% identical amino acids, 49 positions total). Repeated sequences are depicted by arrows.



the sequences of all HTLV-III strains examined to date.

There is also similarity between three subsequences of the PrP 27-30 protein and the carboxyl-terminus of the HTLV-III *pol* polyprotein (Fig. 2). This similarity is shared with the 3'-portion of the *pol* polyprotein of visna virus, a retrovirus with a genomic organization that is similar to HTLV-III and which induces a slow progressive central nervous system disorder in sheep⁸. These protein products are predicted to consist of an extensive region with β sheet structure flanked by regions with predominantly α helical character (program PRSTRC, Molecular Biology Computer Research Resource). These similarities are not evident in the sequences of other retroviruses.

The DNA sequence similarity we describe could be the result of either conver-

gent or divergent evolution. With regard to the latter possibility, the scrapie agent may be the product of a complete or defective retrovirus genome. The existence of fragments of retrovirus genome in the germ line of vertebrates is commonplace and could explain the similarities between the retrovirus sequence and a conserved sequence found in mammalian cells. Expression of such endogenous sequences differs among differentiated systems, may be triggered by external stimuli and may also vary according to the genotype of the animal (see ref. 9 for review).

The similarity of the genes for scrapie-related PrP 27-30 and the *pol* protein of HTLV-III and visna viruses raises the possibility that the proteins mediate the neurological degeneration associated with these infectious agents. Experiments to test the notion that the 3'-end of the *pol*

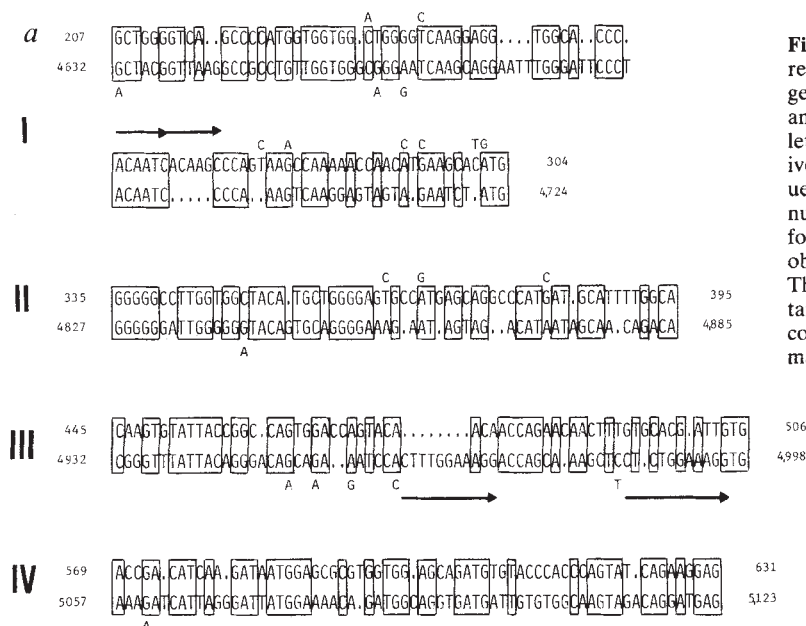


Fig. 1 a, Alignment of the hamster gene for PrP 27-30 scrapie-related gene (above) and the 3'-end of the HTLV-III/LAV *pol* gene (below). Nucleotide changes seen in the mouse PrP 27-30 and in the HTLV-III strain ARV are marked with smaller letters above and below the corresponding sequences, respectively. Deletions of repeats (depicted with arrows) in subsequences I and III are shown. All four subsequences are over 50 nucleotides in length with a similarity above 60%. The scores for the matches range between 14 and 20, well above those obtained when comparing nucleic acid sequences at random. **b**, The significance of the similarities is underscored by the simultaneous presence of the subsequences (I-IV) in an almost colinear fashion. R1-R4 denote direct repeats. The limits of the mature PrP 27-30 are shown.

polyprotein is involved with the neuropathology of AIDS are in progress.

WILLIAM A. HASELTINE

ROBERTO PATARCA

Laboratory of Biochemical
Pharmacology,
Dana-Farber Cancer Institute,
Harvard Medical School,
Boston, Massachusetts 02115, USA

1. Levy, R.M. *et al.* *J. Neurosurg.* **62**, 475-495 (1985).
2. Prusiner, S.B. *et al.* *Cell* **38**, 127-134 (1984).
3. Diringer, H. *et al.* *Nature* **306**, 476-478 (1983).
4. Oesch, B. *et al.* *Cell* **40**, 735-746 (1985).
5. Goldwater, P.N. *et al.* *Lancet* **ii**, 447-448 (1985).
6. Smith, T.F.J. *J. molec. Biol.* **147**, 195-197 (1981).
7. Ratner, L. *et al.* *Nature* **313**, 277-284 (1985).
8. Sonigo, P. *et al.* *Cell* **42**, 369-382 (1985).
9. Weiss, R., Teich, N., Varmus, H. & Coffin, J. (eds) *RNA Tumor Viruses*, (Cold Spring Harbor Laboratory, New York, 1982).

Siberian fire as "nuclear winter" guide

SIR—That the capacity of global models to predict the future can be well tested by their capacity to reconstruct past events is generally agreed¹, as is the definition of normal winter as the numerical equivalent of $>5 \times 10^3$ degree-days (with the degrees in Fahrenheit). One-dimensional radiative-convective models of the interaction of radiation, aerosols and the atmosphere have recently been used² to predict cooling of about 10^4 degree (F)-days in response to the injection of some 100×10^{12} g of smoke, much of which might arise from some 10^5 to 10^6 km² of wildfires, themselves caused by a major nuclear exchange. More recent three-dimensional interactive global climate models³, however, predict only some 10^2 degree-days of cooling from between 20 to 180×10^{12} g of realistic "grey" smoke even in the worst case, at midsummer.

The new models assume smoke injection on a continental scale followed by global transport. Temperature drops are from 6 to 18°C, associated with optical depths between 1 and 3, but decay to ambient conditions in between two weeks and a month. These results agree in large part with the observed outcome of a real smoke injection event of similar magnitude and greater persistence — the great Siberian fire in July and August, 1915.

In a series of wildfires described as "gigantic", some 10^5 to 10^6 km² of Siberia burned (see refs 5,6). The meteorological impact was significant throughout the region (Fig. 1). Reductions of hours of direct sunlight averaged 16 per cent in July and 35 per cent in August, with some stations (for example, Ust Kalskoye, 53°N, 91°E) reporting⁷ more than a 50 per cent reduction. A continuous region the size of Germany, the 250,000 km² lying between the Angara and lower Tunguska rivers, was "completely devastated"⁸. On a conservative estimate, the amount of biomass consumed (including peat, grassland and standing timber) must have produced

Similarity of *mas* and rhodopsin gene products

SIR—Recently a new human cellular oncogene *mas* encoding an integral membrane protein with seven potential transmembrane domains has been described¹. Despite the close similarity in hydrophobic profiles between *mas* and rhodopsin², a protein of the visual system that also has seven potential transmembrane domains, no sequence homology between the two proteins was detected. However, using our method of computer-assisted comparison of amino-acid sequences³, we have found a marked homology between bovine rhodopsin and the *mas* protein and a weak homology between the β -adrenergic receptor⁴ and *mas* proteins.

Alignment of *mas* and bovine rhodopsin is shown in Fig. 1 below. Counting gaps

as substitutions regardless of their length, the sequences exhibit 25% homology. The probability that this is a chance occurrence is less than 3.0×10^{-5} . Of the 48 amino-acids that are invariant among the eight members of the rhodopsin family sequenced to date⁵, 24 are conserved (identical or chemically similar) in *mas*.

HIDENORI HAYASHIDA

KEI-ICHI KUMA

TAKASHI MIYATA

Department of Biology,
Faculty of Science,
Kyushu University,
Fukuoka 812, Japan

1. Young, D., Waitches, G., Birchmeier, C., Fasano, O. & Wigler, M. *Cell* **45**, 711-719 (1986).
2. Nathans, J. & Hogness, D.S. *Cell* **34**, 807-814 (1983).
3. Toh, H., Hayashida, H. & Miyata, T. *Nature* **305**, 827-829 (1983).
4. Dixon, R.A. *et al.* *Nature* **321**, 75-79 (1986).
5. Nathans, J., Thomas, D. & Hogness, D.S. *Science* **232**, 193-202 (1986).

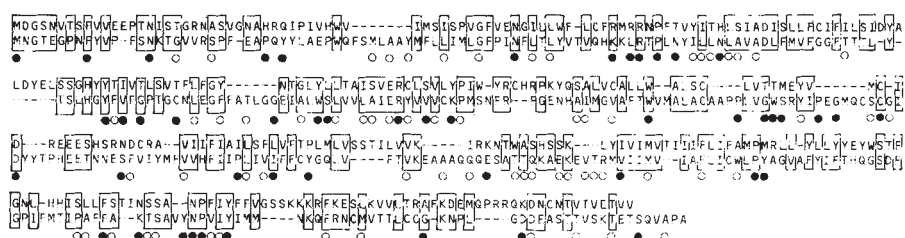


Fig. 1 Alignment of amino acid sequences of *mas* (top) and bovine rhodopsin. Amino acids that are identical or conservative changes (chemically similar) are boxed. Gaps (—) are inserted to increase sequence similarity. The positions of identical (•) and chemically similar (○) amino acids among the eight known sequences of rhodopsin are shown.

more than the 20×10^{12} g of smoke taken as a lower limit in the more recent models. Estimating the amount of smoke produced by widespread fires is notoriously difficult, as the variety of estimates of smoke produced in nuclear exchanges has shown, but central Siberia's rich biotreme (it carries the largest boreal forest in the world) suggests there is a clear possibility that the smoke produced may have equalled the 180×10^{12} g of the worst case³.

It is similarly difficult to assess the contemporary accounts of the 1915 fires. When observers reported that upwards of 1 million km² of forest had been destroyed, they were dealing with an area less than 20 per cent of which had been surveyed⁹.

But even if estimates of smoke production are confined to the 1.4 to 1.6×10^{15} km² of surveyed (Class IV density) mature coniferous timber close enough to rivers for economic harvesting the outcome could well have been 20×10^{12} g of smoke. The unsurveyed remainder, even with the warning against exaggeration offered by reference 10, could easily yield some 100×10^{12} g.

The principal investigators on the scene in 1915, V.B. Shostakovitch, A. Vosnesensky and J. Belyaev of the Irkutsk Magnetic-Meteorology Observatory, fortunately despatched 500 questionnaires throughout Siberia⁷. The 350 respon-

dents, including 142 meteorologists, defined the limits in which fires had occurred as between 70°N (corresponding to the tree line) and 52°N, and between longitudes 60° and 112°, a total area of about 4 million km².

In the area of about 1.8 million km² in which the densest smoke (which was co-extensive with the fires) was reported, visibility at times fell to between 4 and 20 m. Reductions of visibility below 100 m occurred over an area of more than 4 million km² and in places more than 1,500 km from the fire zone. These phenomena persisted for an average of 51 days, from which it is possible to estimate the smoke mass at 40×10^{12} g if the cloud were 1 km thick or 50×10^{12} g if it were 2 km thick¹¹.

The fires were precipitated by a phenomenal drought, the worst since records had first been kept in 1870. According to Furgae¹², the desiccation was so extreme that virtually all the black (coniferous) taiga was in a combustible state, which is borne out by contemporary meteorological records from which modern flammability indices may be derived^{13,14}. Many of the fires which occurred were "topping fires" in which flames first engulf the high branches of trees; smoke from such a fire at Tillamook (Oregon) in 1933 rose to 12 km and the sky was dark at noon.

As well as trees, the fire burned the thick humus of the Yeleni larch forest east

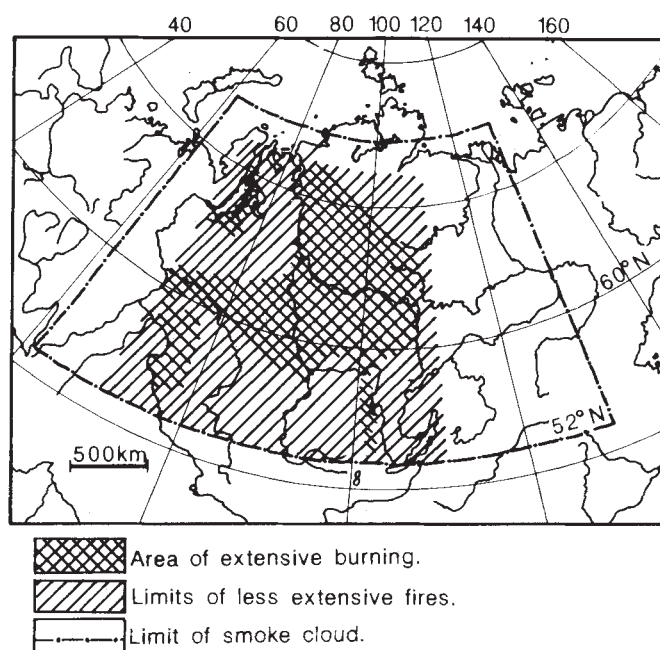


Fig. 1 The areas affected by the Siberian wildfire of 1915. Data from refs 4, 8, 11.

of the Yenesei. The extensive accumulation of biomass over 500,000 km² of peat in central Siberia dried out in many places and burned to a depth of 2 m. Thus there were large injections of between 10¹³ and 10¹⁴ g of smoke from smouldering fires and injections of the same scale at higher altitudes from the burning trees. Even so, the smoke did not appear outside the continent of Asia in contrast with the 1950 fire in Alberta, Canada, which produced optical effects as far away as Britain.

Nevertheless, the Siberian fire of 1915 produced a number of regional effects. High altitude smoke attenuated the solar flux reaching the boundary layer, producing transient cooling of the order of 10°C, although low-altitude smoke with optical depth of less than 5 contributed to the heating of the boundary layer which led to the high temperatures observed locally.

During the densest smoke, one station recorded a week of temperatures no less than 8°C below the monthly trend for August, accompanied by a suppression of the magnitude of the diurnal temperature variation to as little as 2°C. Even so, hard frost was conspicuously absent from the smoke zone; Shostakovitch¹⁴ reports only one night at the end of August when the temperature fell 2°C below zero — hardly a rare circumstance in Siberia.

The agricultural consequences of the smoke were not as great as would be expected on some models. While there are anecdotal reports of smaller harvests, the chief effect⁷ seems to have been to retard the time of harvest by 10 to 15 days (despite the expectation that early ripening in that "dry and hot season" would have brought forward the harvest). It should also be noted that the fire aerosols of 1915 were dense enough to deposit soot and

other combustion products which are said to have sickened cattle and to have killed some hunters in the *taiga*. Yet, even after a prolonged drought, the loss of insolation and the transient cooling does not appear to have brought a catastrophic crop stress in the region.

Obviously there are differences between the pattern of smoke injection over Siberia in 1915 and that postulated in the atmospheric models predicting nuclear winter after a nuclear exchange, but there are also grounds for believing that the differences are not critical. Thus nuclear winter is calculated to be most severe at high latitudes in mid-summer¹⁷, which description accommodates the Siberian fire. The fact that the smoke remained concentrated over Siberia and did not spread out into a circumglobal belt means that the concentration of nuclear winter effects should have been all the greater there, which should also have compensated for the gradual injection of the combustion products into the lower atmosphere.

The quantities of combustion products from the Siberian fire are indeed, not much less than the quantities expected by the climatic modellers after a nuclear exchange, for which reason it would have been expected that the fire would have produced global effects. In reality, the contemporary records show some regional effects, transient (10 to 30 day) cooling (by 5 to 15°C) in response to optical depths of about 3, which are phenomena of the kind associated with nuclear winter. Yet the Siberian fire and its effects differ markedly from what would be expected from the scenario of nuclear winter — the persisting inhomogeneity of the smoke cloud and its failure to leave the boundary layer.

Because the scale of the Siberian fire of 1915 is so similar to that of the conflagrations predicted by some in the aftermath of nuclear war, much could be done to further our understanding of the predictions of the climatic models by studies in Siberia. For example, tree-ring studies of surviving trees in the region could throw light on the course of climatic change in central Siberia in 1915 and after. As a natural event, the Siberian fire is unique in scale, and is unlikely ever to be repeated.

RUSSELL SEITZ

Center for International Affairs,
Harvard University,
Cambridge,
Massachusetts 02138, USA

1. Proc. Int. Conf. on Nuc. War, 452 (Erice, 1984).
2. Turco, R.P., Toon, O.B., Ackerman, T.P., Pollack, J.B. & Sagan, C. *Science* **222**, 1283–92 (1983).
3. Thompson, S., Ramaswamy, V. & Covey, C. *J. geophys. Res.* (in the press).
4. Sverizhev, V.M. et al. in *Ecological and Pedographic Consequences of Nuclear War*, pp.65–66 (Computer Centre, Academy of Sciences, Moscow, 1985).
5. *Forest Industry of the Soviet Union. 1917–1957*, 175 (Directorate of Timber, Moscow, 1958).
6. Pittock, A.B. et al. in *Environmental Consequences of Nuclear War*, vol. 1, 359 (Wiley, Chichester, 1986).
7. Shostakovitch, V.B., *Ocherki Zemledeliu Vostochnoi Sibiri* **47**, 1–9 (1924).
8. Shostakovitch, V.B. *J. Forestry* **23**, 365–71 (1925).
9. Baievsky, B. in *Forest Resources of Siberia*, (Trade Information Bulletin no. 378, Bureau of Foreign and Domestic Commerce, Washington, 1925).
10. Kurbatski, N.P. *Problema Leshnykh Pozharov in Voznikovenie Leshnykh Pozharov*, (Moscow, 1964).
11. Ginsburg, A.S. & Golitsyn, G.S. in *Proc. 2nd All-Union Conf. on Problems of Peace and Prevention of Nuclear War* (Moscow, 1986).
12. Furgaev, V. pp.273–320 in *Problems of Forest Pyrology* (Krasnoyarsk, 1970).
13. Meshcherskaya, A.V. & Blazhevich, V.G., *Tr. Gl. Geofiz.* **400**, 118 (1977).
14. Shostakovitch, V.B. *Records of Irkutsk Observatory 1915* (Irkutsk, 1917).
15. Covey, C., Schneider, S.H. & Thompson, S.L., *Nature* **308**, 21–25 (1984).
16. T. Malone et al. in A.B. Pittock et al., *Environmental Consequences of Nuclear War* vol: 1, Ch. 5, p. 359. SCOPE 28 (Wiley, Chichester, 1986.)
17. Jones, P.D., Wigley, T.M.L. & Wright, P.B. *Nature* **332**, 433 (1986).

Testis size and hermaphroditism

SIR—A recent article in *News and Views*¹ drew attention to an item of major ethnic variation in the field of human reproductive biology. Testis size and rates of dizygotic twinning are markedly lower in Japanese than in Europeans, while the highest rates of dizygotic twinning have been reported in African blacks. These differences might solve another puzzle — the observed differences in the chromosome constitution in human hermaphroditism.

Patients with true hermaphroditism have both testicular and ovarian tissue. Their most common sex chromosome constitution is XX, accounting for about 60 per cent of reported cases² and common among African blacks. By contrast, XY chromosomes have been found in only 15 per cent of hermaphroditism, yet this seems to be the most common type in Japan². The ethnic differences in chromosome constitution of patients with hermaphro-

ditism could be related to variations in testis size.

Several lines of evidence suggest that during embryogenesis, the mammalian gonad may need to grow to a certain size by a given stage of development in order to differentiate into a testis, failing which the gonad becomes an ovary³. In populations with small testes, a few XY individuals may fail to reach the threshold, leading to some ovarian development. On the other hand in populations with large gonads some testicular development may occur in a few XX individuals.

In both XX and XY hermaphrodites, an ovary tends to be situated on the left side and a testis or ovotestis on the right. The bilateral asymmetry of the gonads in true hermaphroditism can be related to an asymmetry of growth in human fetal gonads³. It is noteworthy therefore that the same asymmetry is seen in the testes of adults in three population samples¹, thus providing additional evidence of a relationship between testis size and human hermaphroditism.

URSULA MITTWOCH

Department of Genetics and Biometry,
Wolfson House, 4 Stephenson Way,
University College London,
London NW1 2HE, UK

1. Diamond, J.M. *Nature* 320, 488-489 (1986).

2. Simpson, J.L. in *Sex Differentiation and Chromosomal Abnormalities* (eds Summitt, R.L. & Bergsma, D.) 9-35 (Liss, New York, 1978).

3. Mittwoch, U. *Ann. hum. Genet.* 50, 103-121 (1986).

Accuracy of testing for Huntington's disease

SIR—Earlier this year, *Nature* published two letters^{1,2} giving opposing views on the use of the recently discovered G8 marker for presymptomatic diagnosis of Huntington's disease (HD). We indicated that the "at risk" individual should be provided the opportunity to make an informed choice as to whether he or she wished to be tested but that it was not appropriate to proceed with presymptomatic testing until the question of non-allelic heterogeneity had been adequately addressed. Until further studies being carried out in several centres had been completed, we could not adequately define the accuracy of the presymptomatic test, so that it was impossible to provide the data necessary for an informed choice. Any results would carry a degree of uncertainty that could not at that time be estimated by the genetic counsellor on the basis of existing data. (We were not arguing for a greater level of accuracy in testing, as was suggested³, but simply that the level of accuracy being quoted should have a firm basis.)

Linkage studies of large HD pedigrees with the G8 probe have since produced enough data to estimate accurately the possibility of non-allelic heterogeneity. The data, which will be presented at the

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a rather scientific character. They need not arise out of anything published in *Nature*, but those that do should not be highly technical comments on Articles or Letters (where the Matters Arising section remains appropriate). □

International congress of Human Genetics in Berlin next month, indicate that if a second locus does cause the disease in some families, it is present in at most 10 per cent of HD pedigrees. Using this figure in assessing genetic risk, it is now scientifically justifiable to proceed with presymptomatic diagnosis.

In our previous letter we also suggested that presymptomatic diagnosis in HD is not comparable to most other medical tests. Healthy "at risk" individuals will be taking a test that will predict their own future, one perhaps dominated by a devastating untreatable disease with both physical and mental effects. The HD gene is known to be associated with psychiatric implications and the risk of suicide is significantly increased. Presymptomatic testing has consequences which may prove damaging both to the individual being tested and to other family members. Clearly the testing should only be performed with considerable counselling and support, both before and after the delivery of the results.

In view of these problems, the Massachusetts General Hospital established the Committee for Presymptomatic Huntington's Disease Testing to discuss the most appropriate method of implementing a testing programme. The committee includes representatives from HD lay organizations, from other institutions and from the New England HD Center, as well as experts in medicolegal issues. The committee members indicated that, as with other new medical procedures, the most responsible course of action would be to begin presymptomatic HD testing as a controlled clinical trial. This would permit investigators to assess the most effective means of delivering the information while closely monitoring and exploring methods to deal with, and possibly prevent, negative consequences. The results obtained would also have implications in many other late-onset neurological and behavioural disorders for which similar linkage tests may soon be developed.

Experimental testing programmes will begin in the autumn of 1986 at the Massachusetts General Hospital, Johns Hopkins University and Columbia University and are also planned at several other centers. We are making the G8 probe available through the Committee for Presymptomatic Huntington's Disease Testing to

any other institutions interested in initiating clinical testing programmes approved by their local institutional review board, or equivalent ethics committee. We believe that this approach will result in a clearer understanding of both the potential positive and negative consequences of presymptomatic HD diagnosis by linkage analysis and will also be in the best interest of "at risk" individuals and their families.

JAMES F. GUSELLA

Neurogenetics Laboratory,
Massachusetts General Hospital,
Boston, Massachusetts 02114, USA

1. Watt, D.C., Lindenbaum, R.H., Jonasson, J.A. & Edwards, J.H. *Nature* 320, 21 (1986).

2. Gusella, J.F. *Nature* 320, 21-22 (1986).

3. Maddox, J. *Nature* 320, 11 (1986).

Faster than the eye can see

SIR—In their otherwise lucid observations of a long-duration lightning flash Baker and Baker-Blocker¹ say that a 400 ms flash "is too brief to allow interpretation, since the duration of a perceptual experience is around 0.4 s for the dark-adapted visual system. . . ." But this evades the question of what they mean by "interpretation" and "perceptual experience" because two or more transient flashes, clicks, or taps separated by ≥ 30 ms (that is at > 30 Hz) are easily detected as an ordered temporal sequence. For mixed sensory modalities, for example click and tap, this interval may be ≥ 50 ms, especially when the somatosensory stimulus precedes the auditory one². Surely such psychophysical judgements, by whatever method obtained, involve "interpretation" (discrimination of the physical parameters of each transient) over intervals briefer than 400 ms.

But I agree that transients repeating faster than about 5-7 Hz are hard to count verbally and harder still to replicate by motor output, and I take it that this fact somehow lies behind your correspondents' statement. Also, the long duration (≥ 500 ms) of electrocortical field potential resonances and of relevant multiple neural firings following any suprathreshold, impulse-like sensory stimulus is well known from studies on both humans and nonhumans; less well known and accepted is the evidence for the absence of a human report of tactile sensation if the late resonances of a somatosensory evoked potential are experimentally aborted by neurosurgical manoeuvres³.

HILTON STOWELL

ERBP Laboratory,
120 Nature Creek,
Milledgeville,
Georgia 31061, USA

1. Baker, D.G. & Baker-Blocker, A. *Nature* 322, 215 (1986).

2. Stowell, H. in *Neuropsychol. & Cognit.*, Vol. 2: NATO ASI, MCG, Augusta GA, USA 1980 (eds. Maltaesha, R.N. & Hartlage, L.C.) 565-589 (Nijhoff, The Hague, 1982).

3. Libet, B., Wright, E.W., Feinstein, B. & Pearl, D.K. *Brain* 102, 193-224 (1979).

Remaking the Universe

Laurie M. Brown

The Second Creation: Makers of the Revolution in Twentieth-Century Physics.

By Robert P. Crease and Charles C. Mann.
Macmillan, New York: 1986. Pp. 480. \$25.

THE serious popularization of science is a difficult art, one that requires a broad grasp of the scientific enterprise considered and a discriminating sense of its history. Included in the latter are intellectual development, social setting, principal actors and their backgrounds and personalities. Its authors should combine a strong pedagogical urge with a talent for writing readable prose, and they should be able to simplify, without oversimplifying. In most respects, *The Second Creation* is a satisfying popularization of what may be the most glamorous and philosophically nourishing part of this century's physics (though by no means the part of greatest economic, social or political importance). Unfortunately, this otherwise admirable work is marred by some incorrect and misleading scientific statements, all the more distressing as they could have been caught by any competent physics student. The deeper parts of the subject fare better than the rudiments, perhaps explaining why the errors were not pointed out by the eminent scientists who, according to the authors, "read and commented, sometimes copiously, upon early drafts of chapters".

Robert Crease, who is a graduate student in philosophy at Columbia University, and Charles Mann, who is a writer on science and technology, set out to describe the growth of elementary particle physics, emphasizing recent developments and providing minimal background in the ideas of quantum mechanics. Original citations are in evidence, but the book is based mainly on a limited number of secondary sources; it relies heavily on recent conferences organized by physicists to study the history of their subject. The authors attended some of these conferences and also visited the leading workplaces, taking advantage of these occasions to interview 125 physicists and to obtain a few unpublished documents.

Before enlarging on the positive side of this book, let me indicate the kind of error that I referred to in the first paragraph. In the discussion of Planck's quantum hypothesis (p. 24), it is stated that the oscillators are restricted to discrete frequencies; they are not, it is their energies that are quantized. Again, the first example of a matrix that is shown (on p. 49) is not a matrix at all: the "numbers" 1 and 2 appearing in the two by two array actually represent Child 1 and Child 2,

that is the *names* of children. (No rule is given for the multiplication of such names!) A new way of obtaining relativistic electrons is proposed (p. 81): "What happens when a flashlight beam hits a wall? The electrons in the wall absorb and emit photons, in the process moving at the enormous velocities where relativity is



On his bike — Niels Bohr, with his wife, photographed by George Gamow. The picture is taken from Gamow's *Thirty Years That Shook Physics: The Story of Quantum Theory*, which was first published in 1966 and has now been reissued by Dover, New York/Constable, London.

important." In discussing the famous tau-theta puzzle of the 1950s, which led to the discovery of the breakdown of parity conservation in weak interactions (p. 204), the authors explain: "A stationary pion has a parity of -1 . Ignoring a few additional factors [my emphasis], the two-pion decay has a total parity of $-1 \times -1 = +1$. The three pion decay has a parity of $-1 \times -1 \times -1 = -1$. In this way the tau and theta were distinguished."

In view of such blatant nonsense, how can I call this a satisfying popularization?

(And what could possibly justify the enthusiastic blurbs of distinguished physicists on the book jacket?) For one thing, the density of howlers of the type I have quoted is not high, less than one per ten pages. For another, the book is fun to read: its colourful journalistic style, the enthusiasm of the scientists whose interviews make up the major, and by far the most valuable part of the text, and the intrinsic drama of the creation of the so-called Standard Model (the present, perhaps only temporary, candidate for "the end of physics") make the book difficult to put down. As a writer on the history of physics, I can be only envious of the way the authors evoke a particular scientist's personality in a brief paragraph or two.

In comparing *The Second Creation* with a serious and scholarly history that was greatly praised in these pages (321, 123) not long ago, Abraham Pais's *Inward Bound* (only Part I is called a "history" by Pais, Part II being a "memoir"), one must assume that they are addressed to rather different audiences. Pais writes down many equations and emphasizes technical points that are essentially inaccessible to anyone without an advanced degree in physics. Crease and Mann err in the opposite direction: some of their arguments, as I have tried to show, are too simplistic to be meaningful. Perhaps there is no solution to the dilemma of pedagogy versus readability. The knowledgeable reader will simply ignore the "explanations" of things that he knows better, and the novice will have to glean what he can of the intellectual structure, while getting a sense of the excitement of the chase, the conflict of personality and the rest of the human drama.

Regarding the quality of the history gained from the interviews, it is probably comparable to, though not as high as that in, say, the famous *Archives for the History of Quantum Physics* (mostly the work of Thomas Kuhn and collaborators). In the latter case, the interviewers prepared themselves carefully in advance, and often challenged the physicist being interviewed or confronted him with his own earlier accounts. One does not get the impression that the present authors were anything but sympathetic and accommodating to the views presented by their scientific informants, who seemed only too glad to engage in philosophically rambling (and at times, perhaps, self-serving) accounts of their achievements, and of their brilliant, but unpublished, anticipatory hunches. Almost all the legends and myths of particle physics are repeated, often in several voices, which shows that they are either really true or truly myth, I suppose. □

Laurie M. Brown is Professor of Physics and Astronomy, Northwestern University, Evanston, Illinois 60201, USA.

Working together in the wetlands

G.J. Wainwright

Sweet Track to Glastonbury: The Somerset Levels in Prehistory. By Bryony and John Coles. *Thames & Hudson*:1986. Pp.200. £18.

THE research and educational value of archaeology lies in its multidisciplinary character; the subject draws upon specialists from the biological and physical sciences and brings them together with those whose primary concern is the study of past human cultures. Nowhere is this fusion better exemplified than in the study of wetlands; in the past two decades, this has developed into one of the most pressing areas of international research.

In this book — the twelfth in a series dealing with advances in our understanding of the human past — Bryony and John Coles describe their pioneering work in the Somerset Levels of south-west Britain. The field and publication programme has been in progress since 1963, through the Coles' Somerset Levels Project. It is justly regarded as being very influential in its research directions and methodology, and has served as a model for research in the wetlands of Europe and North America. The book does justice not only to the importance of the subject but to the high reputation of its authors as innovators and managers of the complex multidisciplinary studies that wetland studies generate.

The core of the book describes the recognition and subsequent work upon the prehistoric wooden trackways, including the Sweet Track — the oldest road in the world. These trackways were built from before 3000 BC and connected the small islands of higher ground which projected above the marsh. Investigation of them has provided a new understanding of early human society, its organization and economic base, and the sophisticated fashion in which the people of the time managed the available natural resources through manipulation of their environment. Because of the very nature of wetlands, organic remains are preserved — sometimes these are spectacular, such as the bog burials of northern Europe. More relevant to this project are the palaeoenvironmental studies of pollen, macro-fossils, coleoptera, diatoms and molluscs; the involvement of dendrochronological and radiocarbon laboratories; and the conservation of waterlogged timbers over 5,000 years old.

Apart from containing thorough accounts of these various facets of the project, the book also conveys the message that, as a result of land recla-

mation schemes and exploitation for minerals, the wetlands of the world are a diminishing scientific asset. This is now recognized by English Heritage which supports a number of projects to rescue the evidence before it is destroyed. Moreover, the clear message to emerge is that this response, although worthy, is inadequate: a representative sample of wetlands should also be conserved for the enjoyment and instruction of future generations. The range of interests represented by them is so broad that only the establishment of a Wetland Resource Centre will suffice to mobilize the multidisciplinary expertise necessary for such an enterprise. The particular value of this attractive and compelling book is that it will bring this need to the attention of a wide public. In all, one hopes that it will lay the foundation for a cooperative venture on an international scale, and both authors and publisher are to be congratulated on their enterprise. □

G.J. Wainwright is Principal Inspector at the Historic Buildings and Monuments Commission for England, 23 Savile Row, London W1X 2HE, UK.

HB one hundred

A. Robert Neurath

Hepatitis B. Edited by Robert J. Gerety. *Academic*:1985. Pp.470. \$87, £72.50.

THE first epidemic of hepatitis B (HB) recognizable as such, occurred about 100 years ago. Only 80 years later, in 1965, did the discovery of Australia antigen (=HB surface antigen; HBsAg) and its association with HB provide the basis for scientific attack on the disease. Failure to find an *in vitro* tissue culture system for replication of the HB virus (HBV) then retarded progress in unravelling its biology until 1979, when the problem was in part overcome by the successful cloning of HBV DNA and the expression of HBV proteins in eukaryotic and prokaryotic cells transfected with HBV rDNA.

In this book, our accumulated knowledge of all aspects of HB is comprehensively and succinctly summarized in 18 chapters contributed by 19 authors. The editor characterizes the endeavour as an "attempt to write a *definitive* book on the subject". Considering the vigour of research on HB and the number of outstanding questions concerning the molecular biology of HBV, the immunopathogenesis, therapy and prophylaxis of HB, and the association of HB with primary hepatocellular carcinoma, such an attempt is not entirely realistic. This limitation is recognized by most of the authors, who devote adequate space to unresolved issues and future prospects.

The book is not without its defects, among them an unnecessary degree of repetition of material. There is, too, other evidence of lack of editorial firmness in that a number of contradictory statements are left hanging. This is acceptable when the differences are subjective (to drink or not to drink alcohol when infected with HBV), but disturbing when scientific data are described: according to Table II in Chapter 4, small HB e antigen (HBeAg) has a molecular weight of about 100 kD and is associated with albumin; but in Chapter 14 we learn that the molecular weight is 35 kD (with an incorrect reference).

The second and third chapters, describing HBsAg and HBcAg, respectively, would have benefited from more detailed consideration of the variety of immunoassays for these two antigens and the corresponding antibodies. Moreover, the characterization of HBsAg and HBcAg, dealt with again in Chapters 14, 17 and 18, is imprecise — the fact that saccharides are present in HBsAg is not mentioned; the data on its polypeptide composition are outdated. The N-terminal 140 amino acids of the HBcAg polypeptide do not have a protamine-like structure. In fact, the C-terminal portion of HBcAg has such properties. The competitive RIA for anti-HBc most commonly used utilizes ¹²⁵I-anti-HBc rather than ¹²⁵I-HBcAg (p.39).

Chapter 4 is a comprehensive report on HBeAg. However, the authors seem not to have been fully aware of certain related work carried out outside their laboratory. They also wonder about the origin of "e" in HBeAg — it corresponds, in fact, to the first letter of the second author's name (Espmark) in the paper describing the discovery of HBeAg. The hypothesis that glutaraldehyde-polymerized albumin mediates the attachment of HBV to hepatocytes is untenable.

The chapters describing HB vaccines do not give recognition to the finding that distinct vaccines differ in efficiency when administered to immunocompromised recipients.

Notwithstanding these shortcomings, and the fact that little attention is given to genetic factors and to immunoregulatory mechanisms influencing the immune response to HBV, the book must be counted a success. Researchers in a wide variety of disciplines contribute to unravelling the mystery of hepatitis B. All of them — virologists, epidemiologists, immunologists, physicians and oncologists, not to mention those in public health organizations, blood banks and producers of vaccines, blood derivatives and diagnostic reagents — will find it helpful both for reference and as background reading. □

A. Robert Neurath is Head of the Laboratory of Biochemical Virology, The Lindsley F. Kimball Research Institute of The New York Blood Center, New York, New York 10021, USA.

The bird brained and beyond

Helmut V.B. Hirsch

Memory, Imprinting, and the Brain: An Inquiry into Mechanisms. By Gabriel Horn. Clarendon: 1986. Pp.315. Hbk £25, \$44; pbk £15, \$24.95.

HERE is a highly personal chronicle of the efforts of one group of investigators to understand information storage in the brain. First comes localization of the sites at which information is stored, then the attempt to understand the function of these sites and their relationship to other parts of the brain. The ultimate goal is to study the cellular mechanisms of information storage. Twelve chapters describe this approach, taking the reader step by step through a series of studies on habituation and on imprinting in birds. An appendix giving appropriate details is provided for those unfamiliar with bird neuroanatomy.

The author demonstrates a quality rare in today's world of scientific specialization. He has set himself the task of studying a very important biological problem without allowing himself to be boxed in by the methods at his command, or the borders of recognized fields. Where necessary he sought the help of others and has consistently designed interesting and informative experiments. No one experiment is "critical", but a progressive series of them leads us quite a way towards understanding information storage in the nervous system.

We see in this book how a research programme can grow and develop along with techniques in the field. In early experiments, labelled amino acids were used to localize crudely areas in the chick brain in which protein synthesis, and thus possibly activity and growth, is high during imprinting. Subsequently, autoradiography was used to home in on the exact location of areas involved in information storage. Thus, technical advances coupled with experimental ingenuity were crucial in advancing understanding of imprinting and of the underlying process of information storage.

The first, and perhaps the weakest, chapter of the book deals with the decrement in neuronal and behavioural responses upon repeated presentation of the same stimulus. Habituation is compared with other forms of learning that involve elaboration of a new response or modification (not simply loss) of an old one. Imprinting in birds is put forth as a model for studying these forms of information storage; this is the focus of the remainder of the book.

Incorporation of radioactive lysine or

radioactive uracil is used to identify brain regions at which there are high levels of activity and/or growth during imprinting. One hypothesis is that these are potential sites for information storage during imprinting, but there are alternative hypotheses. For example, incorporation of lysine or uracil at these sites may reflect non-specific side-effects of training, such as differences in activity level between imprinted and control animals. By combining careful behavioural observations with studies of incorporation of uracil, Horn has been able to rule out many of the alternative hypotheses.

Autoradiographic techniques increase the precision in identifying potential sites for information storage during imprinting; a restricted part of the forebrain, the "intermediate and medial part of the hyperstriatum ventrale" (IMHV), can be identified with this procedure. The author would, however, probably be the first to admit that in order to test critically the hypothesis that the IMHV is the area that stores information during imprinting, manipulative neural experiments must be performed.

Some aspects of the imprinting process involve areas outside the IMHV. Chicks will imprint on many different types of target, but they appear to have a predisposition to prefer certain "biologically appropriate" objects such as a stuffed, rotating jungle fowl. This predisposition is probably determined intrinsically, because it does not require previous exposure to the appropriate target, but can be activated by exposure to a lighted environment or running in an activity wheel in the dark; it is not abolished by lesioning the IMHV.

There is further evidence that neural systems concerned with preferences for biologically appropriate imprinting targets are to some extent distinct from those concerned with storage of the information needed to identify particular objects as familiar. A reduction in the level of noradrenaline in the brain affects imprinting to targets such as a rotating, flashing red box, but not to the jungle fowl. In contrast, testosterone levels in the plasma are correlated with imprinting to the jungle fowl, but not to the red box. A preference for the jungle fowl may therefore reflect a special-purpose information storage system, one affected by testosterone levels; learning to recognize a particular object as familiar may involve a second, general-purpose mechanism, one modulated by the level of noradrenaline. Under natural conditions imprinting probably takes place at an age when the predisposition for natural stimuli is fully developed and ensures that the chick will select an appropriate imprinting target. Continued exposure to a particular target allows the chick to store information needed to recognize it as familiar. Thus the chick

can distinguish its mother from other females of the species.

There are some flaws in the book, mostly cosmetic. There are far too many typographical errors; in several places parts of sentences are missing. Also, figure 5.3 is mislabelled, and several figures lack error bars. It is a great shame that minor irritants such as these were allowed to intrude into what is otherwise a most impressive book.

Horn's monograph speaks to several audiences. It gives students of animal behaviour an impetus to consider the neuronal basis of behaviour, and it shows neuroscientists the power of carefully designed behavioural studies. Indeed, the book is a reminder that a very basic goal in neuroscience is understanding the role of the nervous system in behaviour. Finally, and perhaps most important, in the same fine tradition as K.D. Roeder's *Nerve Cells and Insect Behavior* it will inspire students to take up the challenge posed by the nervous system. The author has prepared a *vade-mecum*; taking us step by step along his path, he presents a highly personal account — but without blowing his own horn. □

Helmut V.B. Hirsch is Professor in the Neurobiology Research Center, State University of New York at Albany, Albany, New York 12222, USA.

High Quality Products
from FUNAKOSHI



Recombinant
Mouse-
Interferon-β

Blue-Cotton

—mutagen
adsorbent—



rIFN is Pyrogen Free and highly purified material (>99%), prepared from Recombinant E. coli by affinity chromatographies described by S. Matsuda, et al (to be published in J. Interferon Res). Stable at -20°C.

Blue-cotton is adsorbent for mutagens containing quite low concentration in aqueous solutions, such as Urine, Feces, Foods, River water, etc. And the adsorbed mutagens are easily eluted by ammonia-methanol solution.

FUNAKOSHI PHARMACEUTICAL CO., LTD.
Science Business Div. Research Material Sect.
2-3 Surugadai, Kanda, Chiyoda-ku, Tokyo, Japan
Telephone : Tokyo 03-293-2352
Telefax : 81-3-295-5545
Telex : J28489 FUNA

Plenum: The Natural Choice GENE TRANSFER

edited by **Raju Kucherlapati**

Gene Transfer deals with many of the methods used to transfer genetic information into mammalian cells and describes a number of parasexual systems.

0-306-42325-1/441 pp. + index/ill./1986
\$59.50 (\$71.40 outside US & Canada)
text adoption price on orders of six or more
copies: \$35.00

CONTRACEPTIVE STERIODS

Pharmacology and Safety

edited by **A. T. Gregoire** and
Richard P. Blye

This new volume considers a broad spectrum of topics concerned with oral contraceptive steroids including their adverse effects and benefits in clinical use. A volume in the series *Reproductive Biology*.

0-306-42393-6/proceedings/382 pp. + index
ill./1986/\$65.00 (\$78.00 outside US & Canada)

GENETIC DISORDERS AND THE FETUS

**Diagnosis, Prevention, and
Treatment • Second Edition**

edited by **Aubrey Milunsky**

From reviews of the first edition:

"...thought provoking...excellent summaries..."

— *American Journal of Human Genetics*

"...a major text in the field of prenatal medicine and genetic disorders of the human fetus.... This text is strongly recommended, for specific reference information, as a potential text, and as a worthwhile review for future American Board of Medical Genetics examinations."

— *American Journal of Medical Genetics*

0-306-42301-4/879 pp. + index/ill./1986
\$89.50 (\$107.40 outside US & Canada)

NMR IN THE LIFE SCIENCES

edited by **E. Morton Bradbury**
and **Claudio Nicolini**

The importance of *in vivo* nuclear magnetic resonance (NMR) and NMR imaging in biomedical science and medical diagnosis has resulted in an explosion of growth in many major areas of NMR research. This book presents seventeen papers that bring the reader up to date with recent advances in the field. Volume 107 in the NATO ASI Series: Series A: Life Sciences.

0-306-42279-4/proceedings/246 pp./ill.
1986/\$45.00 (\$54.00 outside US & Canada)

Plenum Publishing Corporation
233 Spring Street
New York, N.Y. 10013-1578



In the United Kingdom:
88/90 Middlesex Street
London E1 7EZ, England

Cooperation in the abstract

Athel Cornish-Bowden

**Cooperativity Theory in Biochemistry:
Steady-State and Equilibrium Systems.** By
Terrell L. Hill. Springer-Verlag: 1985.
Pp. 459. DM 398.

COOPERATIVITY is of central importance for the control of biochemical processes. Although most enzymes obey rate equations in which the rate of reaction shows a less than proportional dependence on the concentration of the substrate, and although this is quite satisfactory for many purposes, it does not allow for the possibility of a rapid switching from one metabolic state to another in response to a primary stimulus that may be quite small. As a result, certain proteins, not all of them enzymes, have evolved the property of cooperativity, whereby a large change in behaviour can result from a relatively small change in conditions.

The classic example is provided by the binding of oxygen to haemoglobin. Here, a cooperative response is needed to allow most of the binding sites to be filled when blood passes through the lungs and most of them to be vacated before it returns, despite a relatively small difference between the partial pressure of oxygen in the lungs and that in the tissues. This phenomenon has been known since the beginning of the century, and study of it has contributed in large part to our present understanding of cooperativity, not least because haemoglobin, a binding protein and not an enzyme, can properly be treated by equilibrium methods. When used to study enzyme cooperativity, by contrast, equilibrium theory is at best an approximation.

So far as most biochemists are concerned, the main developments in theories of cooperativity occurred about 20 years ago, with the publication in 1965 of the "allosteric model" of Monod, Wyman and Changeux, and in 1966 of the "sequential model" of Koshland, Némethy and Filmer. These models have dominated discussion ever since, despite the fact that they are both "quasi-equilibrium models" that seek to explain kinetic phenomena as if they did no more than mirror behaviour that one might observe if the system were at equilibrium. The allosteric model, in particular, has had wide intuitive appeal, although its central postulate of conformational symmetry has no foundation in physical principles.

Nonetheless, even if for many biochemists the field ceased to be interesting in about 1970, there have been a few people who have tried to place the proper-

ties of cooperative systems in biochemistry on a firmer basis of physics. One of the most notable of them has been Terrell Hill, who has spent some years showing that in addition to the traditional empirical approach to cooperativity it is possible to work from a much more fundamental molecular point of view, considering systems in terms of partition functions and, indeed, all of the concepts of statistical thermodynamics. As the author mentions in the preface to *Cooperativity Theory in Biochemistry*, his new book can be regarded as a continuation of his successful textbook *Introduction to Statistical Thermodynamics*, published in 1960. More specifically, it gathers together the material published by the author and a few collaborators between 1977 and 1982, in the hope that this will bring it to the attention of a wider audience.

For physicists already familiar with statistical thermodynamics, this book will provide a valuable introduction to the kinds of problem that exist in the biological sciences, and which will eventually need to be understood on a much more fundamental level. Biochemists on the other hand, despite the appeal of the title, will find that it is extremely demanding and makes few concessions to any lack of familiarity with statistical mechanics. The treatment throughout is abstract and mathematical, with little discussion of experimental systems. The most familiar examples — haemoglobin, aspartate carbamoyltransferase and so on — are barely mentioned. Instead, the main experimental system considered is cooperativity in muscle proteins, a very important example, certainly, but one that is rather too complicated to be approached as a first step.

The treatment presented in the book is very much the author's own creation, developed largely in isolation from others interested in similar problems. A high proportion of the references, both in the book and in his original papers, are to his own work. Hill's contribution to the field is widely recognized, though less widely studied, and to have it summarized in a single book is very useful. □

Athel Cornish-Bowden, a Lecturer in the Department of Biochemistry, University of Birmingham B15 2TT, UK, is currently on leave of absence at the Universidad de Chile, Santiago, Chile.

New in paperback

- *In the Name of Eugenics: Genetics and the Uses of Human Heredity*, by Daniel J. Kevles. Publisher is Penguin/California University Press, price is £5.95, \$9.95. For review see *Nature* 317, 481 (1985).
- *The Making of Mind: A Personal Account of Soviet Psychology*, by A.R. Luria. Publisher is Harvard University Press, price is \$7.95, £6.75. For review see *Nature* 283, 797 (1980).

Implications of a two-component marble-cake mantle

Claude J. Allègre* & Donald L. Turcotte†

* Laboratoire de Géochimie et Cosmochimie, Institut de Physique du Globe de Paris, 4, place Jussieu, 75230 Paris Cedex 05, France

† Department of Geological Sciences, Cornell University, Ithaca, New York 14853-1504, USA

It is suggested that the upper mantle contains elongated strips of subducted oceanic lithosphere. These strips are stretched and thinned by the normal and shear strains in the convecting mantle, and are destroyed by being reprocessed at ocean ridges or, on the centimetre scale, by dissolution processes; the result is a marble-cake mantle. Simple theoretical calculations, together with isotopic and structural observations made on high-temperature peridotite massifs, lead to a comprehensive marble-cake model which is consistent with most isotopic and mechanical constraints.

GEOCHEMICAL studies have shown that the upper mantle reservoir is depleted in incompatible elements and is complementary to the enriched continental crustal reservoir¹. We first outline how the geodynamic cycle is responsible for this chemical fractionation and how it leads to a marble-cake mantle. The cycle consists of three stages:

(1) Mantle rock ascends passively beneath an ocean ridge in response to seafloor spreading, and becomes partially molten due to the decrease in pressure on ascent. The magma percolates through the residual solid and then solidifies to form the oceanic crust, with an average thickness of ~6 km. The result is a two-layer structure for the rigid ocean lithosphere. The upper part of the lithosphere is the solidified magma and the lower part is the complementary residual solid. The oceanic crust is itself layered, with near-surface basalts overlying gabbro and dolerite that were produced by fractional crystallization. The residual solid also has a vertical stratification. The uppermost mantle rock is highly depleted in the low-melting-temperature component, and grades into undepleted mantle over a depth range of ~50 km.

(2) The oceanic lithosphere spreads, more or less symmetrically, away from the ocean ridge. A substantial fraction of the oceanic crust is chemically altered by seawater interactions, particularly near the spreading centre where temperatures are high and extensive hydrothermal circulations occur². The sea floor is also coated by the deposition of sediments. Eventually the oceanic lithosphere plunges back into the mantle at a subduction zone. Seafloor sediments are incorporated into the descending lithosphere and parts of the overriding plate may also be incorporated. At a depth of ~125 km the descending oceanic crust is heated sufficiently to melt; the resulting magma ascends to the surface and forms island arc volcanoes. This volcanism is responsible for the creation of a large fraction of the continental crust. During this process part of the oceanic crust is 'scavenged' of incompatible elements. Through the double process of oceanic crust formation and island arc volcanism, the original mantle rock is depleted in incompatible elements before the rock is returned to the mantle at the subduction zone. These incompatible elements are incorporated into the enriched continental crust.

(3) The cold, layered oceanic lithosphere sinks through the upper mantle to the base of the convecting layer, at either 670 km for layered mantle convection or 2,800 km for (less probable) whole mantle convection. In either case heat is transferred from below by conduction. The cool lithosphere is heated and softened on a timescale of 50 Myr. The lithosphere then enters the convecting region of the mantle and is deformed. The primary deformation process is one of elongation of the layered lithosphere³⁻⁹. The recently subducted lithosphere is mixed back

into the surrounding mantle, a substantial fraction of which is composed of oceanic lithosphere that was subducted at earlier times. Thus, a large fraction of the mantle will be composed of layered oceanic lithosphere which has been extensively deformed but not totally destroyed and homogenized. Hofmann and Hart¹⁰ have shown that atomic diffusion plays a role in the homogenization of the mantle only on scales of a metre or less, because the solid-state diffusion coefficient is so small. Values of the relevant diffusion coefficients are estimated to be in the range $D = 10^{-14}$ – 10^{-16} cm² s⁻¹. Over the age of the Earth, 4.5×10^9 yr, the corresponding range of diffusion lengths is 0.1–10 m. In contrast, diffusion coefficients in magmas are much larger ($\sim 10^{-7}$ cm² s⁻¹), so that melting is a very efficient way to achieve chemical and isotopic homogenization. We conclude that the subducted lithosphere is mixed back into the mantle by convection, but that diffusive mixing is significant only on small scales. This process is known as 'streamline' mixing and has been extensively studied in polymer science⁹. The mantle is composed of a matrix of discrete, elongated layers of subducted oceanic lithosphere. Each layer has its own isotopic, chemical and age identity. The older the layer the more it will have been elongated; on average, the older layers will be thinner. The mantle thus has the appearance of a marble cake¹¹.

The marble cake comprises the enriched oceanic crust which has been partially depleted by subduction zone volcanism, and the complementary, highly depleted upper mantle. As they convect through the mantle these materials are affected by phase changes, such as the basalt-eclogite, olivine-spinel and pyroxene-garnet transformations. At the base of the upper mantle the marble cake is a garnierite-spinel mixture rather than an eclogite-peridotite mixture; however, for convenience, we will retain the eclogite-peridotite description for the entire process. This is a chemical rather than a mineralogical classification.

Two processes act to destroy the thinning, layered oceanic lithosphere. The first is the reprocessing of the mantle at an ocean ridge. This destroys the previously subducted ocean crust by melting. The second process is the dissolution of the thinned eclogitic strips by a complex sequence of metamorphic processes including recrystallization and solid-state diffusion.

Studies in experimental petrology confirm that the basalts of the oceanic crust can be obtained by melting either eclogite or fertile peridotite⁴⁷. Melting of a marble-cake mantle explains the isotopic heterogeneities of mid-ocean-ridge basalts (MORB). The melting of a marble-cake mantle homogenizes a strongly heterogeneous mantle to various degrees, depending on the extent of melting. For example, MORB from the East Pacific Rise is nearly homogeneous. As the spreading rate is high on the East Pacific Rise, large volumes of mantle are partially

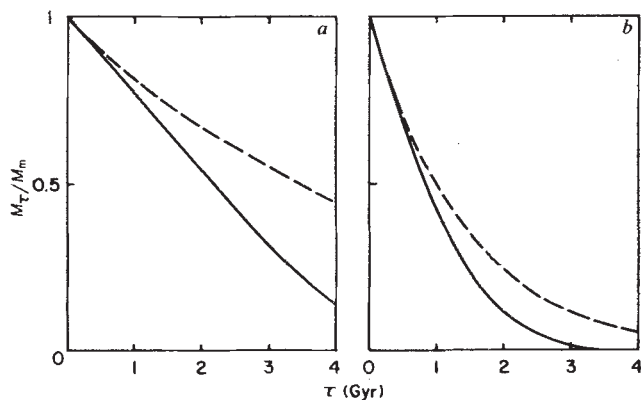


Fig. 1 The fraction of the mantle older than τ , M'_τ/M_m , plotted as a function of time τ before the present. *a*, Whole mantle convection ($\tau_p = 4.96$ Gyr); *b*, layered mantle convection ($\tau_p = 1.28$ Gyr). Dashed lines are for a constant processing rate, solid lines for an exponentially decreasing processing rate ($\tau_r = 2.5$ Gyr).

melted to form the oceanic crust. On the other hand, seamounts adjacent to the East Pacific Rise have greater variability in isotopic ratios¹². This can be attributed to the much smaller amounts of magma and the localized melting required to form the seamounts.

Mantle recycling

The consequences of the marble-cake model will depend on the rate at which mantle rock is processed by the plate tectonic cycle and the effectiveness of streamline mixing. To simplify the analysis we assume that the rate $\dot{M}$ at which mass is processed into a layered structure at ocean ridges is constant, and that subducted material is uniformly distributed throughout the mantle. We define M_R to be the mass of layered oceanic lithosphere subducted during a time interval of length τ . We can write

$$\frac{dM_R}{d\tau} = \dot{M} - \frac{\dot{M}M_R}{M_m} \quad (1)$$

where M_m is the mass of the mantle which participates in the plate tectonic convective cycle—the whole mantle for whole mantle convection, and the upper mantle for layered mantle convection. The first term on the right-hand side of equation (1) represents the addition of subducted, layered oceanic lithosphere and the second term is the loss of this material by reprocessing at an ocean ridge. Integrating with the boundary condition $M_R = 0$ at $\tau = 0$ we obtain

$$M_R = M_m \left\{ 1 - \exp \left[-\frac{\dot{M}\tau}{M_m} \right] \right\} \quad (2)$$

where $\tau_p \equiv M_m/\dot{M}$ is the characteristic time for processing the mantle in the plate tectonic cycle.

The processing rate $\dot{M}$ is given by

$$\dot{M} = \rho_m h_p \frac{dS}{d\tau} \quad (3)$$

where ρ_m is the mean density and h_p the thickness of the layered structure and $dS/d\tau$ is the rate at which plate area is created (or subducted). Taking $dS/d\tau = 1.35 \times 10^3 \text{ cm}^2 \text{ s}^{-1}$, $h_p = 60 \text{ km}$ and $\rho_m = 3.2 \text{ g cm}^{-3}$, we obtain $\dot{M} = 2.6 \times 10^{10} \text{ g s}^{-1}$. For layered mantle convection the characteristic time for processing the mantle is $\tau_p = 1.28$ Gyr; for whole mantle convection $\tau_p = 4.96$ Gyr. The mass of the mantle processed before time τ , $M'_\tau(\tau)$, is given by $M'_\tau = M_m - M_R$. The fraction of the mantle processed before time τ , M'_τ/M_m , is shown in Fig. 1 as a function of τ (with τ measured from the present), for whole mantle and layered mantle convection (dashed lines in Fig. 1*a* and *b*, respectively).

A similar calculation can be carried out assuming a time-dependent rate of seafloor spreading. Instead of taking $\dot{M}/M_m$ to be constant we take $\dot{M}/M_m = (1/\tau_p) e^{\tau/\tau_r}$, where τ_r is the time

measured backward from the present. The assumption is that the rate of seafloor spreading decreases exponentially with time constant τ_r , as the radioactive heat sources within the Earth decay. Taking $\tau_r = 2.5$ Gyr and the values of τ_p used above, the fraction of the mantle older than τ is plotted as a function of τ in Fig. 1 (solid lines). Comparison of the constant-rate process (dashed lines) with the time-dependent process (solid lines) shows that the latter is more efficient in eliminating old and pristine mantle. However, for simplicity we will consider below only the constant-rate process.

Streamline mixing

We now turn to the problem of layer stretching. As stated above, we hypothesize that the subducted oceanic crust becomes entrained in the convecting mantle, and is deformed by the strains associated with thermal convection. A number of papers have dealt with this problem. Richter and Ribe³ considered a simple two-dimensional model of thermal convection and determined the thinning of a heterogeneity with time; Richter *et al.*⁴ showed that time-dependent convection is more efficient in thinning than steady convection. Olson *et al.*^{5,6} studied both convective and diffusive dispersion, and pointed out the relative importance of shear and normal strains on thinning. Hoffman and McKenzie⁷ carried out a series of numerical calculations of layer stretching and applied their results to the stretching of the subducted oceanic lithosphere. Numerical calculations of streamline mixing have also been carried out by Gurnis and Davies⁸, who argue that, with reprocessing at ocean ridges, the process is relatively inefficient at mixing the mantle.

A much more extensive literature exists on layer stretching in the polymer sciences, and this has been reviewed by Ottino and Chella⁹. Extensive experiments have been carried out and a variety of convective systems have been studied analytically and numerically. If shear strains dominate the convective mixing, then the layer thinning in the time τ is given approximately by

$$\frac{\delta}{\delta_0} = \frac{1}{1 + \dot{\epsilon}\tau} \quad (4)$$

where δ is the layer thickness, δ_0 is the initial layer thickness at $\tau = 0$ and $\dot{\epsilon}$ is the rate of strain.

If normal strains dominate the convective mixing, then the layer thinning is given approximately by

$$\frac{\delta}{\delta_0} = \exp[-\dot{\epsilon}\tau] \quad (5)$$

Normal strains are more effective in layer thinning than shear strains. The exponential form, equation (5), was favoured by Hoffman and McKenzie⁷, based on numerical calculations; we will use this form for our comparisons.

The element of layered structure that is subducted in the

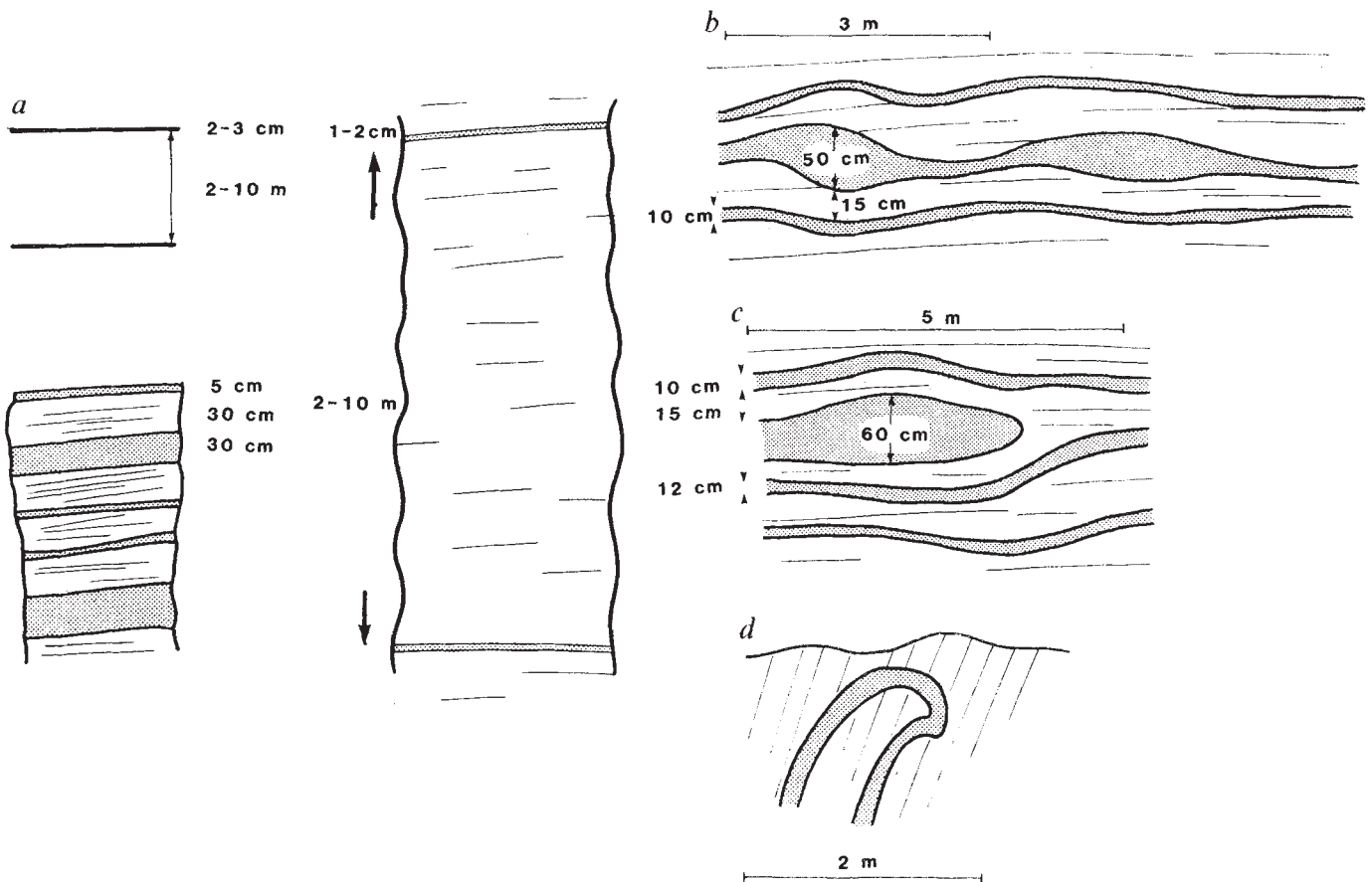


Fig. 2 Occurrences of pyroxenite layers in the Beni Bousera high-temperature peridotite. Grey, pyroxenite; white, lherzolite with foliation. *a*, Occurrences in an outcrop with no folding; *b-d*, occurrences with folding and boudinage.

interval $d\tau$ is $\dot{M} d\tau$. The fraction of this mass that has not been reprocessed in the time τ is

$$dM'_i = \dot{M} \exp \left[-\frac{\dot{M}}{M_m} \tau d\tau \right] \quad (6)$$

Using equation (5), the area of lithosphere, dS'_τ , that was subducted in the interval $d\tau$ and stretched for a time τ is given by

$$dS'_\tau = \frac{dM'_\tau}{\rho_m \delta} = \frac{\dot{M}}{\rho_m \delta_0} \exp \left[\left(\dot{\epsilon} - \frac{\dot{M}}{M_m} \right) \tau \right] d\tau \quad (7)$$

and the total area of lithosphere subducted in the time τ is

$$S'_\tau = \int_0^\tau dS'_i = \frac{\dot{M}}{\rho_m \delta_0 (\dot{\epsilon} - \dot{M}/M_m)} \left\{ \exp \left[\left(\dot{\epsilon} - \frac{\dot{M}}{M_m} \right) \tau \right] - 1 \right\} \quad (8)$$

Because the exponential term is large for times of interest it is appropriate to neglect the -1 , after which combining equations (5) and (8) gives

$$\log S'_\tau = B - \left(1 - \frac{\dot{M}}{\dot{\epsilon} M_m} \right) \log \delta \quad (9)$$

where

$$B = \log \left\{ \frac{\dot{M}}{\rho_m \delta_0 (\dot{\epsilon} - \dot{M}/M_m)} \right\} + \left(1 - \frac{\dot{M}}{\dot{\epsilon} M_m} \right) \log \delta_0 \quad (10)$$

We will use equation (9) to make comparisons with observations.

Note that in the subduction process, the lithosphere may be broken into several pieces of unknown dimensions. Thus, the layers of centimetre size in the mantle may be the result of stretching bodies with an original thickness of <6 km.

High-temperature peridotite massifs

We have derived a model for the streamline mixing of subducted oceanic lithosphere into the mantle; for observational evidence supporting this model, we now turn to pieces of the Earth's mantle which have been transported to the surface, and which display a marble-cake structure. These exposed samples are the high-temperature peridotites, also called orogenic lherzolite massifs, such as Ronda in Spain¹³, Beni Bousera in Morocco¹⁴, Lanzo in the Alps^{15,16}, Lherz in the Pyrenees¹⁷ and Tinaquillo in Venezuela¹⁸.

These massifs are composed principally of a lherzolite matrix with layers of pyroxenite. The layers represent a few per cent of the massifs, they are basaltic in bulk composition, and their thicknesses vary from a few centimetres to a few metres. Their lengths can be mapped over distances of several hundred metres. They are parallel to the foliation of the massif and clearly exhibit plastic deformation, including folding and boudinage, and mineral segregation for the smallest layers^{14,16}.

Chemical studies have shown that the lherzolites are not pyrolytic in composition but have been partially or totally depleted in clinopyroxene. The degree of depletion is variable from place to place. Trace element studies have confirmed that these lherzolites display a clear rare-earth element (REE) depletion pattern¹⁹⁻²⁴.

The pyroxenite layers are extremely variable in mineralogical composition; their complexities have recently been deciphered^{22,23} by quantitative REE modelling. These layers record a complex history with several episodes of melting; many of them have compositions close to what would be expected for a residual, altered oceanic crust after the extraction of magma at a subduction zone²².

Isotopic studies have added important information. Strontium and neodymium isotopes have shown that orogenic lherzolites have the characteristics of the depleted mantle²⁵⁻²⁷. In addition, each massif exhibits a large isotopic variability, almost equivalent to the total variation observed in the oceanic basalt population^{28,29}.

The pyroxenites are statistically more radiogenic in strontium than the peridotites, but no simple isochron can be constructed for each whole massif. Mineral isochrons record the time of tectonic emplacement²⁷; recent surveys of lead and neodymium isotope data have confirmed this interpretation^{28,29}. The extreme isotopic heterogeneity of each massif and the radiogenic character of the pyroxenites can be identified with the oceanic basalt field on isotope correlation diagrams. In addition, it is clear that massifs situated in different geological provinces (Lanzo, Beni Bousera, Lherz) have distinct isotopic signatures²⁸.

For these reasons, it has been proposed that these massifs are representative of the convecting upper mantle, with the pyroxenite layers being recycled oceanic crust^{22,25,28}. Menzies and Murthy³⁰ have argued that the clinopyroxenes in the high-temperature peridotites do not have sufficient K, Rb and Ba to be the source of MORB; however, the distribution of potassium in high-temperature peridotites is extremely heterogeneous. Some have potassium-rich amphiboles, some have potassium on the interstitial boundaries between grains, and some pyroxenites are slightly enriched in potassium. If the degree of partial melting is not too large and if the partial melting samples a reasonably large volume of this heterogeneous source region, it will be possible to provide sufficient potassium to form MORB. Frey *et al.*²⁴ have shown that some peridotites at Ronda have enough potassium to produce basalts, and Boudier and Nicolas¹⁵ have shown the same for Lanzo. In addition, two field observations indicate that orogenic lherzolite can produce basalt by partial melting. In Lanzo, Boudier and Nicolas³¹ observed gabbroic lenses and dykes that were produced *in situ*; in Lherz, the lherzite dykes are clear products of melting of the peridotite, as discussed by Loubet and Allègre²². Both the gabbros and the lherzites are basaltic, the former being tholeiitic and the latter alkalic.

All of this information can be reconciled with a model in which orogenic lherzolites are samples of the convecting mantle, emplaced tectonically at the surface. In order to provide further support for the model, we have carried out a systematic study of gully and trail-cut exposures at Beni Bousera.

The pyroxenite layers have three general characteristics (see Fig. 2). (1) They are usually very thin (2–30 cm), and are dispersed in the lherzolite at intervals of 2–10 m. (2) In some zones the layers occur as sequences interbedded with the lherzolite. In these sequences thick layers (50 cm to 1 m thick) sometimes occur. Pyroxenite layers locally represent 10–15% of the exposure. (3) In some cases, extensive deformation of the layers has occurred. Deformation includes folding, as previously described in ref. 14, isoclinal folds and, less frequently, boudinage or telescoping of two layers. In some zones the small layers terminate in cusp-like structures due to the segregation of pyroxene. All of these features are consistent with intense plastic deformation over a long period of time.

The statistics of layer thicknesses have been determined, in order to provide quantitative information on the convecting mantle. The number of layers N_c with a thickness greater than δ is plotted as a function of δ in Fig. 3. Assuming that the Beni Bousera outcrop studied is representative of the mantle, and that each layer extends across the outcrop, we can write equation (9) as:

$$\log N_c = B' - \left(1 - \frac{\dot{M}}{\dot{M}_m}\right) \log \delta \quad (11)$$

The data shown in Fig. 3 correlate well with a slope slightly shallower than -1 (~ -0.85 to -0.90); thus, we conclude that $[(\dot{M}/\dot{M}_m)/\dot{\epsilon}] = 0.10$ – 0.15 . Taking $\tau_p = \dot{M}_m/\dot{M} = 1.38$ Gyr for layered mantle convection we find that $\dot{\epsilon} = (1.5$ – $2.3) \times 10^{-16} \text{ s}^{-1}$;

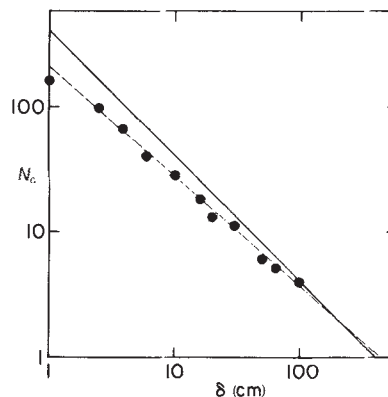


Fig. 3 Number of pyroxenite layers exposed at Beni Bousera with a thickness greater than δ , N_c , as a function of δ . Points, observations; dashed line, $N_c \propto \delta^{-0.87}$; solid line, $N_c \propto \delta^{-1}$.

taking $\tau_p = 4.96$ Gyr for whole mantle convection we find that $\dot{\epsilon} = (4.3$ – $6.4) \times 10^{-17} \text{ s}^{-1}$. These strain rates are about an order of magnitude lower than typical values associated with mantle convection⁴⁵ ($\sim 10^{-15} \text{ s}^{-1}$). This difference can be attributed to the presence of both normal and shear strains in the mantle. As pointed out above, normal strains are much more effective than shear strains in stretching the layers. As only a fraction of the total strain field is normal strain, the effective stretching strain will be less than the total strain present.

The pyroxenite layers can be homogenized with the mantle by dissolution processes. Although the observed distribution of layer thicknesses is in good agreement with the stretching hypothesis in the size range 2–100 cm, there is a systematic divergence for thicknesses ≤ 2 cm: the number of layers thinner than ~ 2 cm is much less than predicted. We attribute the loss of thin layers to dissolution processes that homogenize and destroy thin layers. These dissolution processes are likely to include both recrystallization and diffusion. As a measure of these processes, we use the diffusion relation

$$\delta = (D\tau)^{1/2} \quad (12)$$

Taking $\delta = 2$ cm and $\tau = 10^8$ yr we find $D = 10^{-15} \text{ cm}^2 \text{ s}^{-1}$, which is a reasonable value for solid-state diffusion in the mantle¹⁰.

An independent measure of the role of dissolution in homogenization is the volumetric fraction occupied by the pyroxenite layers in high-temperature peridotites. Boudier and Nicolas¹⁵ estimate 1.5% for Lanzo, Kornprobst¹⁴ $< 3\%$ for Beni Bousera; our statistics for this massif give 1%. Without homogenization we would expect this ratio to be close to the 10–15% associated with basalt segregation at the ocean ridge. The difference between these two values is evidence that a significant fraction of the marble-cake mantle has been destroyed by homogenization processes.

The relationship between isotopic composition and thickness is more difficult to study because of the complexity of the local situation, the multiplicity of layer compositions, and the occurrence of mantle metamorphism, which has redistributed the elements locally. However, lead isotope studies have shown that rehomogenization does occur between layers at the centimetre scale. Radiogenic lead increases as the thickness of the pyroxenite layers decreases, whereas the original lherzolite is non-radiogenic in lead²⁷.

Discussion

Marble cake and the mesosphere boundary layer. The idea of a marble-cake mantle, as proposed in 1979 by Allègre *et al.*¹¹, was questioned on the basis of rare-gas observations³². Significant quantities of primordial ^3He are found in MORB, whereas prior processing of primordial mantle at an ocean ridge would be expected to remove the primordial rare gases. Thus, the rare-gas data cannot be explained by the marble-cake model³³.

This dilemma is solved by combining the marble-cake mantle model with the mesosphere boundary layer model³⁵. In the latter model, layered mantle convection occurs, with a boundary at the 650-km seismic discontinuity. At the base of the upper convecting layer, a thermal boundary layer develops, due to heat transfer from below. The upper mantle is the depleted, outgassed mantle and the lower mantle is less depleted and less outgassed. Ascending mantle blobs responsible for the generation of ocean island basalts (OIB) result from boundary layer instabilities. The large blobs responsible for Hawaii or Iceland entrain some material from the lower mantle; because the lower mantle is far more enriched in rare gases than the upper mantle, the rare-gas signature of these blobs is relatively primitive. Rare gases can also diffuse across this boundary: because helium has a much larger diffusion coefficient than the heavy rare gases, the ³He enrichment of the upper mantle can be attributed to diffusion³⁶.

The difference in isotopic composition between MORB and OIB is more statistical than systematic. If we consider the global results, including the Indian Ocean ridges and the Pacific islands, the fields of MORB and OIB have a large overlap³⁷. However, the statistical differences must be explained. For most OIB the isotopic composition can be explained by the preferential melting of eclogitic strips relative to the peridotitic matrix, whereas MORB is closer to (but not so depleted as) the ultramafic matrix. Several effects probably contribute to this difference. (1) The fact that the OIB are statistically more alkali-rich than MORB suggests that more eclogite (either in the strips or as residual isotopic halos in the peridotite) is involved in the OIB melt. (2) The smaller volumes of OIB volcanoes retain the mantle heterogeneities, whereas the large volumes of melt associated with MORB result in homogenization. (3) Because eclogite is denser than peridotite, large pieces of the oceanic crust would have the tendency to concentrate at the base of the upper mantle. Thus, OIB should have more garnet and clinopyroxene in their source than average peridotites³⁸.

The Dupal anomaly^{39,40} seems to be created by a different phenomenon, since equivalent isotopic patterns have not been observed in high-temperature peridotites. The interpretation given previously⁴¹, that this global pattern of isotopic anomalies is created by old continental lithosphere delaminated several hundred million years ago, remains plausible. This hypothesis, proposed in ref. 41 for the origin of OIB in general, seems to be true only for the global Dupal anomaly.

The 'mantle isochron'. Mantle rock processed at ocean ridges is isotopically homogenized. On the other hand, the layer stretching due to streamline mixing in the mantle maintains isotopic heterogeneities for thousands of millions of years. Even after the mineralogical signature of a layer is destroyed by dissolution, the local isotopic halo persists. It should be emphasized,

however, that only the extraction of continental crust leads to the depletion of the mantle. Some areas of the mantle are more depleted than others because of differential extraction; however, these differences tend to be erased by mantle convection. This problem has been treated quantitatively by Allègre *et al.*¹¹ using the mean field approximation. They have shown that mantle isochrons are produced by continuous processes and reflect the dynamics of extraction rather than any discontinuous event; these results apply to the marble-cake mantle.

Magmatic inert trace element ratios. Hofmann *et al.*⁴² have studied trace element ratios, such as Nb/U and Ce/Pb, that do not vary appreciably between MORB and OIB but are different from both chondritic and continental crustal values. They inferred two chemical fractionations in the geodynamic cycle: one at the ocean ridges that is purely magmatic, and a second at subduction zones that extracts the continental crust and is much more complex. If we assume that the first process does not fractionate the magmatic inert trace element ratios, OIB and MORB ratios should be similar. If continental extraction does fractionate such ratios, then the observed ratios are consistent with our model.

Physical properties of the mantle. One implication of a marble-cake mantle is seismic anisotropy. Seismic studies have shown that signals propagate more rapidly horizontally than vertically in the upper mantle⁴³⁻⁴⁵. This can be attributed to horizontal layering of two constituents with different seismic velocities. As mantle convection is likely to be dominated by horizontal flows, the layering in a marble-cake mantle is likely to be predominantly horizontal. Olson *et al.*⁵ have argued that a substantial fraction of the observed seismic anisotropy can be attributed to this effect, confirming previous work⁴³ which indicated an even larger effect.

Another implication of the marble-cake hypothesis concerns the rheology of the mantle. Laboratory studies indicate that the eclogitic layers will have a viscosity which is 10²-10³ times larger than the viscosity of the peridotitic matrix. This contrast in viscosities will add a considerable stiffness to the mantle. As the flow will align the layers with the streamline direction, the high-viscosity layers will have relatively little effect on shear flows; however, the stiff layers will strongly inhibit normal strains⁴³. The higher viscosity of the eclogitic layers will also influence the layer stretching associated with streamline mixing. The influence of differential viscosities has not been studied in the geophysical literature³⁻⁸, but has been considered in the polymer literature⁹. Viscosity differentials greater than a factor of three lead to boudinage of the layers. The layers are still stretched and thinned, but at a reduced rate.

We thank Al Hofmann for a very constructive review. This research was partially supported by NSF grant EAR-8518019.

Received 14 April; accepted 14 July 1986.

- Allègre, C. J. *Tectonophysics* **81**, 109-132 (1982).
- Staudigel, H., Hart, S. R. & Richardson, S. H. *Earth planet. Sci. Lett.* **52**, 311 (1981).
- Richter, F. M. & Ribe, N. M. *Earth planet. Sci. Lett.* **43**, 212-222 (1979).
- Richter, F. M., Daly, S. F. & Natof, H. C. *Earth planet. Sci. Lett.* **60**, 178-194 (1982).
- Olson, P., Yuen, D. A. & Balsiger, D. J. *geophys. Res.* **89**, 425-436 (1984).
- Olson, P., Yuen, D. A. & Balsiger, D. *Phys. Earth planet. Inter.* **36**, 291-304 (1984).
- Hofmann, N. R. A. & McKenzie, D. P. *Geophys. J. R. astr. Soc.* **82**, 163-206 (1985).
- Gurnis, M. & Davies, G. F. J. *geophys. Res.* **91**, 6375-6395 (1986).
- Ottino, J. M. & Chella, R. *Polym. Engng Sci.* **23**, 357-379 (1983).
- Hofmann, A. W. & Hart, S. R. *Earth planet. Sci. Lett.* **38**, 44-62 (1978).
- Allègre, C. J., Brévart, O., Dupré, B. & Minster, J. F. *Phil. Trans. R. Soc. A* **297**, 447-477 (1980).
- Zindler, A., Staudigel, H. & Batiza, R. *Earth planet. Sci. Lett.* **70**, 175-195 (1984).
- Dickey, J. *Spec. Pap. miner. Soc. Am.* **3**, 33-49 (1970).
- Kornprobst, J. *Contr. Miner. Petrol.* **23**, 283-322 (1969).
- Boudier, F. & Nicolas, A. *Schweiz. miner. petrogr. Mitt.* **52**, 39-56 (1972).
- Nicolas, A., Bouchez, J. L. & Boudier, F. *Tectonophysics* **14**, 143-171 (1972).
- Conquéré, F. *Bull. Soc. fr. Minér. Cristallogr.* **100**, 42-80; 123-137 (1977).
- Green, D. H. J. *Petrology* **5**, 134-188 (1964).
- Frey, F. A. *Geochim. cosmochim. Acta* **33**, 1429-1447 (1969).
- Loubet, M., Shimizu, N. & Allègre, C. J. *Contr. Miner. Petrol.* **53**, 1-12 (1975).
- Loubet, M. & Allègre, C. J. *Geochim. J.* **13**, 69-75 (1979).
- Loubet, M. & Allègre, C. J. *Nature* **298**, 809-814 (1982).
- Dickey, J. S., Obata, M. & Suen, C. J. *Bull. Ore. Dep. Geol. miner. Ind.* **86**, 79-89 (1977).
- Frey, F. A., Suen, C. J. & Stockman, H. *Geochim. cosmochim. Acta* **49**, 2469-2491 (1985).
- Polvé, M. & Allègre, C. J. *Earth planet. Sci. Lett.* **51**, 71-93 (1980).
- Richard, P. & Allègre, C. J. *Earth planet. Sci. Lett.* **47**, 65-74 (1980).
- Zindler, A., Staudigel, H., Hart, S. R., Endres, R. & Goldstein, S. *Nature* **304**, 226-230 (1983).
- Hamelin, B. & Allègre, C. J. *Eos* **66**, 1114 (1985).
- Reisberg, L. & Zindler, A. *Earth planet. Sci. Lett.* (in the press).
- Menzies, M. & Murthy, V. R. *Earth planet. Sci. Lett.* **38**, 346-354 (1978).
- Boudier, F. & Nicolas, A. *Bull. Ore. St. Dep. Geol. miner. Ind.* **86**, 63-78 (1977).
- Allègre, C. J., Staudacher, Th., Sarda, Ph. & Kurz, M. *Nature* **303**, 762-766 (1983).
- Allègre, C. J., Hart, S. R. & Minster, J. F. *Earth planet. Sci. Lett.* **66**, 177-213 (1983).
- Davies, G. F. J. *geophys. Res.* **89**, 6017-6040 (1984).
- Allègre, C. J. & Turcotte, D. L. *Geophys. Res. Lett.* **12**, 207-210 (1985).
- O'Nions, R. K. & Oxburgh, E. R. *Nature* **306**, 429-431 (1983).
- Hamelin, B., Dupré, B. & Allègre, C. J. *Earth planet. Sci. Lett.* **76**, 288-298 (1986).
- Hofmann, A. W. & White, W. M. *Earth planet. Sci. Lett.* **57**, 421-436 (1982).
- Dupré, B. & Allègre, C. J. *Nature* **303**, 142-146 (1983).
- Hart, S. R. *Nature* **309**, 753-757 (1984).
- McKenzie, D. & O'Nions, R. K. *Nature* **301**, 229-231 (1983).
- Hofmann, A. W., Jochum, K. P., Seufert, M. & White, W. M. *Earth planet. Sci. Lett.* (in the press).
- Anderson, D., J. *geophys. Res.* **66**, 2953-2963 (1961).
- Toksöz, N. & Anderson, D. J. *geophys. Res.* **68**, 1121-1130 (1961).
- Fuchs, K. *Phys. Earth planet. Inter.* **31**, 93-118 (1983).
- Turcotte, D. & Schubert, G. *Geodynamics* (Wiley, New York, 1982).
- Yoder, H. S. *Generation of Basaltic Magma* (National Academy of Sciences, Washington, 1976).

Polarization vision in bees

Samuel Rossel & Rüdiger Wehner

Department of Zoology, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

A solution to the long-standing problem of how honey bees detect polarized light from the sky is presented. The mechanism involves the transformation of polarization information into modulations of perceived brightness while the bee scans the sky by rotating its field of view. Using this technique the bee needs only a very simple strategy to read compass information from the polarization patterns in the sky.

KARL VON FRISCH discovered that honey bees can detect linearly polarized light and can use this ability to derive compass information from the Sun-linked polarization patterns in the sky¹⁻³. Subsequently, the basis of polarization sensitivity in the retina of bees and other invertebrates was shown to lie in the oriented arrays of pigment molecules packed in the membranes of the layered microvillar structure of the rhabdomeric photoreceptors, such that the sensitivity of the receptors is greatest when light is polarized parallel to the microvilli (see ref. 4 for review). Furthermore, it is known for bees that a small, dorsal region of the retina (the POL area) is both essential and sufficient for navigation by means of polarized sky light⁵ and that this specialized retinal area contains an array of polarization-sensitive ultraviolet photoreceptors which equip it for this task⁶⁻⁹. One outstanding problem in the elucidation of polarization vision is how bees use information from the receptor array to analyse and orient by the patterns of polarized sky light.

Here we present a novel and comprehensive solution to this problem by showing how the array of receptors forms a template which the bee uses to scan and match the polarization patterns in the sky. We begin with a brief review of our recent model for the detection of polarized light and go on to report experiments designed to test the model. We conclude that the mechanisms uncovered may be generally operative in polarization-sensitive invertebrates.

Polarized light detection model

Imagine that in each eye of the bee the orientation of microvilli of polarization-sensitive photoreceptors varies in a graded manner through 180°, being horizontal in the anterior and posterior part of the eyes and oriented at all other angles between these positions (Fig. 1). Consider next what happens to the output of such an array when the bee, when presented with a small stationary patch of polarized light, rotates about its vertical body axis. The patch is positioned at 60° above the horizon and stimulates successively a ring of receptors as the bee rotates. The angle between microvilli and the *e*-vector varies gradually during the turn. As a consequence, the output of the array is a modulating signal passing through a peak as the microvilli become parallel to the *e*-vector in the patch. A similar modulation is detected by the bee when the *e*-vector direction is changed, except that the modulation now rises to a peak in a different position on the retina. Thus, we assume that for the analysis of *e*-vector directions in the sky, the bee sweeps its retinal analysers across the *e*-vectors and records which part of the retina is maximally active during the sweep.

This simple model of *e*-vector detection arose from behavioural experiments in which we analysed orientation performance of the bee to small spots of polarized sky light¹⁰⁻¹³. The implications of these experiments are most easily explained if we consider how the scanning mechanism might be used by bees to obtain compass bearings from the polarization patterns in the sky. Consider first a polarization pattern as it is realized

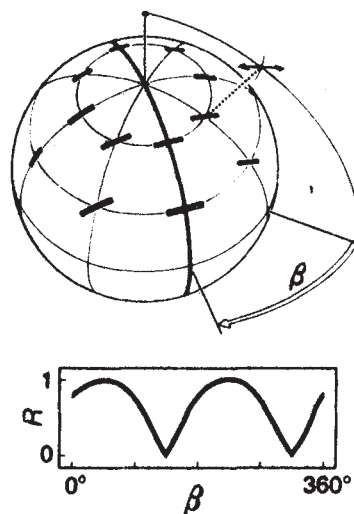


Fig. 1 Mechanism of detecting polarized light by bees. Sphere, visual fields of the left and the right eye (separated by thick continuous lines). The systematic variation of microvillar (analyser) directions of polarization-sensitive photoreceptors is shown for two elevations above the horizon (30 and 60°). Note that the dichroic visual pigments are oriented parallel to the microvilli so the sensitivity of the receptors is greatest when the *e*-vector of polarized light is parallel to the microvilli and least when it is at right angles to them. The operational principles of the model are demonstrated for a situation in which the bee views a patch of polarized light exhibiting a single *e*-vector direction (double arrow). The bee is first aligned with the patch ($\beta = 0^\circ$) and then performs a full turn about its vertical axis ($\beta = 0-360^\circ$), thus sweeping a ring of receptors with different microvilli directions across the *e*-vector. The relative output (*R*, log scale) of successive receptors within the ring is plotted as a function of β . A peak output is induced by the patch of polarized light when its *e*-vector direction is matched by a suitably arranged microvillar direction. The rationale behind the scanning model is that each *e*-vector direction has its own location on the retina, specified by a peak output of a corresponding photoreceptor.

at dawn and dusk (Fig. 2a). When the model bee is aligned with the solar and the antisolar meridian of such a pattern, the layout of polarization analysers over the eyes matches closely the distribution of *e*-vectors in the sky. Given this coincidence, the bee can rotate about its vertical body axis and assume that whenever there is a peak in the output of the *e*-vector detectors, the body is aligned with the solar and the antisolar meridian. It can then turn from this reference direction until it faces the desired compass course.

The situation changes during the day when the Sun is high in the sky (Fig. 2b). In this case there are substantial discrepancies between the *e*-vector pattern and the array of polarization analysers. Thus when the bee, while scanning a small patch of

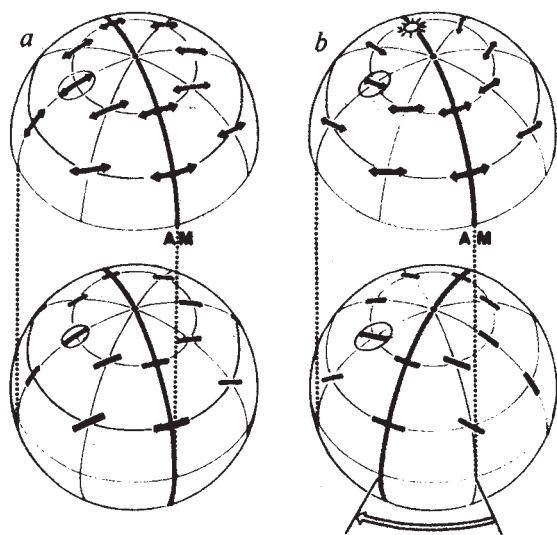


Fig. 2 Use of the scanning model in *e*-vector pattern orientation. **a**, Upper hemisphere, distribution of *e*-vectors (arrows) across the sky when the Sun is at the horizon. The celestial sphere is viewed from the back (AM, antisolar meridian). The pattern of polarized light depends strictly on the elevation of the Sun. As a general rule the angle of polarization (*e*-vector direction) is always perpendicular to the plane containing the Sun, the patch in the sky and the observer in the centre of the celestial hemisphere. The lower sphere shows the layout of analysers within the eyes of the bee. Note that the analysers match the *e*-vectors when the bee is aligned with the solar and the antisolar meridian. Thus, to determine the position of these meridians from one or another *e*-vector in the sky (in the figure exemplified for a vertical *e*-vector) the bee sweeps its field of view across the *e*-vector and assumes that it is aligned with the reference meridians when it detects a peak output from its analyser array. **b**, A different polarization pattern is realized in the sky when the Sun is at an elevation of 65°. When now presented with an isolated *e*-vector (exemplified here by an oblique *e*-vector), the bee again assumes that it is aligned with the solar and the antisolar meridian when it perceives a peak intensity from the patch of polarized light. But, in this case, it is mistaken (open arrow) and ultimately steers an incorrect course.

a daytime pattern, applies its simple method of determining the reference meridians, it may arrive at an incorrect estimate of the position of the meridians. This in turn should lead to errors of orientation expressing the discrepancy between the layout of analysers and the *e*-vector pattern. Our findings show that bees do indeed make consistent orientation errors, depending on the elevation of the Sun and the *e*-vector direction which is viewed by the bee in the sky. Therefore, the interpretation of the orientation data in terms of a scanning mechanism of *e*-vector detection allows us to infer the spatial layout of retinal analysers shown in Fig. 1.

These behavioural interferences are supported by anatomical and optical observations on the POL areas of the bees, which show that the microvillar directions of photoreceptors vary across the retina in the way predicted. However, the details of this correspondence have not been completely checked^{7,9}.

Finally, an important interference derives from combined anatomical and physiological studies, showing that each ommatidium of the POL area contains a set of orthogonally arranged microvillar directions⁷⁻⁹. To accommodate this, we have extended our model by adding a second array of receptors to the one already described (Fig. 3). We then suppose that the visual system compares the outputs of perpendicularly oriented receptors such that the differential response will be maximum when the microvilli of one receptor array are oriented parallel to the *e*-vector and least when the *e*-vector is parallel to the microvilli of the other^{9,14}. Clearly, such antagonistic interactions of crossed analysers enhance the specificity of *e*-vector detection,

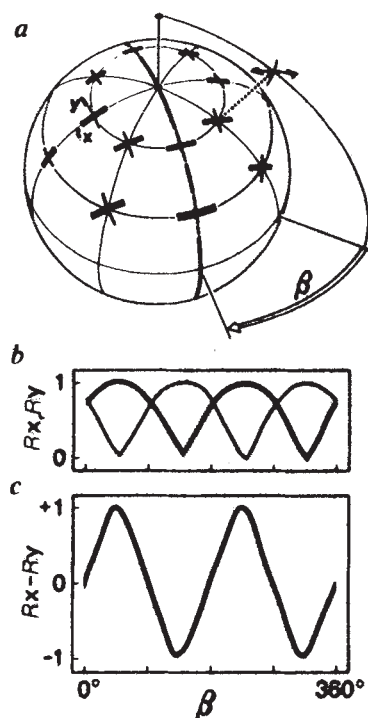


Fig. 3 Extended version of the scanning model. **a**, In addition to the array of receptors shown in Fig. 1 (here called x-type receptors) there is an array of y-type receptors with orthogonally arranged microvillar directions. Equipped with this array of crossed analysers, the bee scans a patch of polarized light displaying a single *e*-vector direction (double arrow). **b**, Output of the x- and y-type receptors (R_x and R_y , log scale) as a function of the horizontal orientation relative to the azimuthal angle β of the patch of polarized light. **c**, Output of the crossed analyser array as a function of β , when the outputs of the x- and y-type receptors are compared differentially by the visual system ($R_x - R_y$, log scale).

but they are not essential in principle for the scanning model to work. In fact, it has already been shown theoretically that only one analyser is needed per ommatidium, if polarized light is to be analysed by a scanning method¹⁵.

Our model can be tested by exploiting the ability of the bee to orient by means of small patches of polarized skylight. As can be deduced from Fig. 2, bees invariably treat a small patch of polarized light of a given *e*-vector as though it were at a fixed azimuthal angle relative to the solar and the antisolar meridian, although this is in fact only true of the pattern of polarization across the sky at dawn and dusk. For instance, bees orient to a patch of light with a horizontal *e*-vector as it lies along the antisolar meridian and to a patch with a vertical *e*-vector as though it is at right angles to the antisolar meridian. Thus we know both where a particular *e*-vector direction is represented within the retina and also what azimuthal angle the bee associates with a patch of light of that *e*-vector direction. The scanning model then predicts that a patch of unpolarized light arranged to evoke a peak output within a particular part of the retina will cause the bee to orient as though the light were polarized in the appropriate direction. Although this prediction is simple, a complex experiment is necessary to test it.

Experimental evidence

The standard method for analysing *e*-vector orientation is to see how polarized light influences the 'waggle dance' bees perform when they return to the hive after foraging. The orientation of this dance on the surface of the comb signals the direction

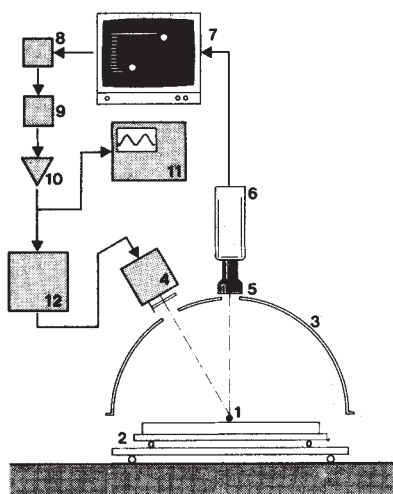


Fig. 4 Experimental device for testing the scanning model. The bee (1) performs its directional dances on horizontal combs (2) covered with a black Plexiglass hemisphere (3). Light from a Xenon arc (4), equipped with an ultraviolet filter (Schott; UG 11), is the only cue for dance orientation. The light stimulus subtends an angle of 10° and is at an elevation of 60° . The horizontal body-rotation of the dancing bee controls the intensity of the Xenon arc. For this purpose two circular pieces (1 mm) of a reflecting foil (3M; 7610) are stuck onto the thorax and the abdomen of the bee, exactly 8 mm apart. The ring lamp (5), which surrounds the lens of a television camera (6), is then switched on and the aperture of the lens closed until the image on the television monitor (7) fades out completely, except for the two light spots reflected from the body of the bee. In each frame the number of lines between the first and the last image signals is counted. Because the distance between the two reflecting spots is constant, the number of lines provides a direct measure of the direction of the medial plane of the bee and hence can be used to control the intensity of the Xenon arc (8, line counter; 9, digital-to-analogue converter; 10, amplifier; 11, storage oscilloscope; 12, power supply of Xenon arc). Maximal and minimal intensities are produced when the bee is aligned either parallel or at right angles, respectively, to the vertical axis of the monitor. With respect to the medial plane of the bee, maximal intensities can be induced at any angle by rotating the camera relative to the azimuthal angle of the Xenon arc. The intensity between maximal and minimal intensity varies by a factor of three. All waggle dances are videotaped and later analysed by measuring the direction of individual waggle runs.

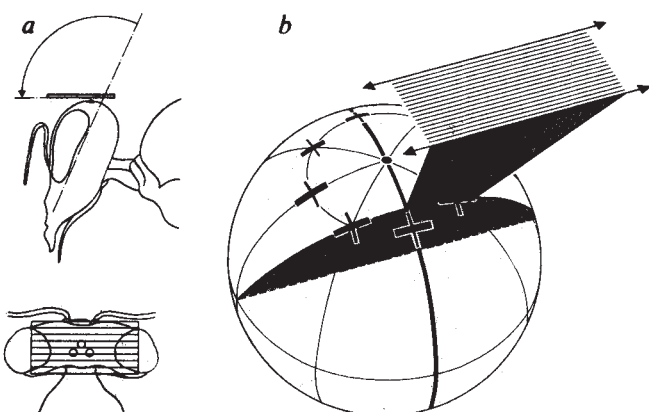


Fig. 5 Method for stimulating the receptors selectively. *a*, Side and dorsal view of the head of the bee that contains a piece of polarizing filter (Polaroid, HNPB). The transmission axis of the filter is parallel to the transverse axis of the head. The inclination of the filter is defined relative to the back plane of the eyes. *b*, Model eyes with crossed analysers (*x*- and *y*-type receptors; Fig. 3) at an elevation of 60° , and part of the overhead filter. Throughout the array the microvillar directions of the *x*-type receptors are defined by the set of planes which include the transverse axis of the eyes (here one plane is shaded). In addition, these planes include the transmission axis of the filter. Therefore, light is polarized by the filter to stimulate predominantly the *x*-type receptors.

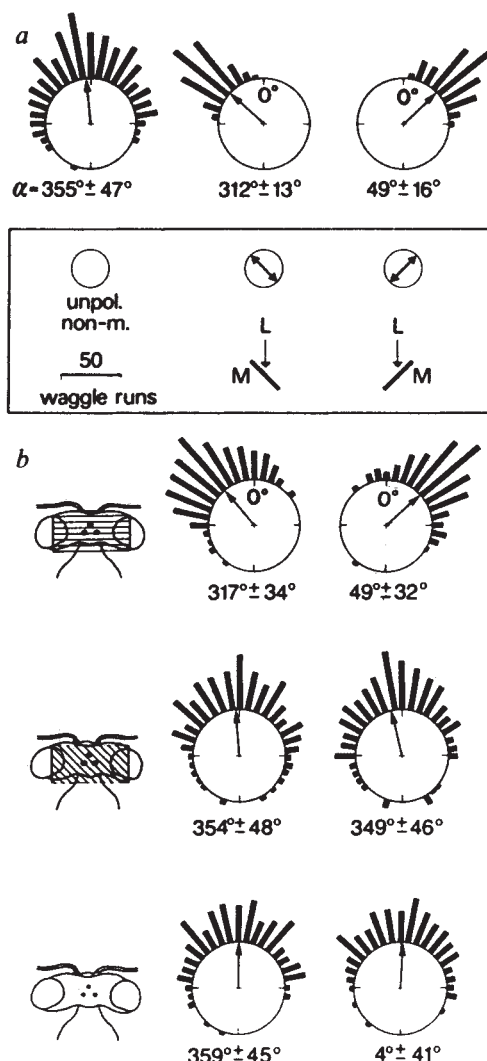


Fig. 6 *a*, Dance directions of bees when presented with a patch of unpolarized non-modulating ultraviolet light and a patch of polarized light with one of two oblique vectors (stimulus conditions are indicated in the upper half of the inset; *e*-vector directions are marked by double arrows). Mean dance directions (α , arrow) are calculated to indicate where the navigating bee expects the light stimulus to occur in the sky (0° marks the antisolar meridian Fig. 2). Data are based on 931 waggle runs recorded from 16 bees. *b*, Dance directions of bees when presented with a patch of unpolarized modulating ultraviolet light. The modulations are arranged to simulate the two *e*-vector directions presented to the bees in *a* (in the inset *M* is the orientation of the bees relative to the azimuthal angle of the light stimulus *L* when its intensity passes through a peak; compare with Fig. 1 for retinal positions of oblique *e*-vectors). In the upper and middle parts of *b* the dancing bees are equipped with an overhead polarization filter (the transmission axis of the filter is oriented parallel and at 45° to the transverse axis of the head). In the lower part of *b* the bees are not carrying overhead filters. Data based on 2,760 waggle runs recorded from 24 bees.

of the foraging site from the hive. On vertical combs in the dark, bees orient with respect to the gravitational vector. On horizontal combs they will orient with respect to the Sun or the polarization patterns across the sky³.

In the present experiments, several individually labelled bees were trained to a feeding station several hundred metres from the hive. The hive had horizontal combs and an artificial light source provided the trained bees with the only directional cue by which they could orient their dance. When given a single patch of polarized light in such a situation, the orientation of

the dance varies with the *e*-vector direction. We sought to mimic this behaviour by means of a beam of unpolarized non-modulating ultraviolet light which is modulated in intensity according to the horizontal orientation of the bee, so that the intensity of the beam peaks when it stimulates a predetermined part of the retina. Details of the servo system are given in Fig. 4.

As we described above, each ommatidium within the POL area is equipped with pairs of orthogonally oriented polarization detectors which we assume interact to give a differential output. At first sight a modulated beam of unpolarized light appears to be an inappropriate stimulus because pairs of crossed analysers will be stimulated equally, so that the differential output will be zero throughout the modulation cycle. To circumvent this effect we selectively stimulated one of the two crossed receptor arrays on its own by covering the eyes with a small piece of polarizing filter oriented to transmit *e*-vectors parallel to the transverse axis of the eyes (Fig. 5). This filter transmits the set of *e*-vector planes which contain lines that are parallel to the transverse axis. Microvilli are oriented over the retinal region of interest so that light transmitted by the filter will stimulate one of the two receptor arrays more than the other. As a control, some bees are equipped with a filter with its transmission axis oriented 45° from the transverse axis, in which case the two crossed analysers should again be stimulated equally.

The light was modulated so that the beam was of peak intensity when it fell on one of two different retinal positions which according to the model contain detectors for one of two oblique *e*-vectors. The orientation of the dances under these conditions is shown in Fig. 6. Bees indeed behave as the scanning model predicts. With the transmission axis of the filter parallel to the transverse axis of the eyes, dances orient as though the bees are presented with one of the two oblique *e*-vector directions. In contrast, with the filter oriented at 45° from the transverse axis of the eyes, the effect disappears and the bees behave as though they are stimulated by a patch of unpolarized, non-modulating ultraviolet light.

Before these results can be considered to provide conclusive evidence in support of our model, we asked whether there is an alternative mechanism of *e*-vector detection which could produce the same results. Previous accounts assumed that each individual ommatidium of the compound eyes is able to assess and report independently on the *e*-vector direction of light falling within its receptive field of view (see refs 3, 16 for review). Thus one might argue that polarization orientation in our experiments is not affected by the intensity modulations of the ultraviolet beam, but is caused by the *e*-vectors projected onto the eyes by the overhead polarization filter. This possibility, however, can be excluded. Because the direction of the *e*-vectors projecting onto the eyes varies as both the bee and filter rotate, different ommatidia would give the bee conflicting and uninterpretable information about *e*-vector directions. Similarly, we would have expected the dance directions to be the same whether the filter was parallel to the transverse axis of the eyes or oriented obliquely to it.

Conclusions

We can now explain the principles underlying polarization vision in bees in terms of the arrangement of dichroic photoreceptors, the visual cues abstracted from these receptors and the influence this processing strategy has on the orientation behaviour.

Specifically, our data show that bees have an array of retinal analysers, each of which is maximally sensitive to a different *e*-vector direction. When a patch of polarized sky light is swept across the array, the perceived direction of its *e*-vector is determined by which analysers within the array produce the largest output. We also deduce from our experiments that *e*-vector

detection by the array of analysers exploits antagonistic interactions between receptors with orthogonally oriented microvilli. We believe that such interactions provide the bee with a high-contrast *e*-vector signal.

The spatial layout of retinal analysers forms the structural basis of *e*-vector detection. In addition, it matches the distribution of *e*-vectors in the sky, when the bee is aligned with the solar and the antisolar meridian (Fig. 2). Thus to determine the position of these reference meridians, the bee turns about its vertical axis and detects when the array of analysers is maximally active.

This simple strategy of reading compass information from the polarization patterns explains the orientation performance of bees whose vision of the sky is restricted to small areas containing a very small range of *e*-vector directions. At dawn and dusk there is a good match between the sky pattern and the array of detectors, but during the day substantial discrepancies occur. Thus, when the bee applies its simple compass strategy it may arrive at an incorrect course, as we have previously described¹⁰⁻¹³. But this shortcoming is not as serious as it might first appear. Under normal circumstances the navigating bee has large areas of the polarization patterns at its disposal. Therefore, it can benefit from the fact that *e*-vectors in each celestial half (defined by the solar and the antisolar meridian) are mirror images of each other (Fig. 2). This implies that error angles induced by *e*-vectors from the left and the right half of the sky are always opposite in sign, and thus cancel each other out^{12,13}. Furthermore, during free flight the navigating bee can resort to powerful backup systems such as the Sun and landmarks^{3,17}.

Finally, note that the structural peculiarities that characterize the array of polarization analysers within the retina of bees seem to be common in invertebrates. The retinas of cephalopods, crustaceans and insects have sets of receptors with orthogonally arranged microvilli (see refs 16, 18 for review). Furthermore, there is increasing evidence that the directions of microvilli vary in fan-shaped arrays across the retina in various groups of insects^{9,19,20}. It seems likely, therefore, that the principles discussed here are generally applicable to invertebrates that detect and use polarized sky light for orientation.

We thank H. Baumann, K. Feller and T. Grob for design and construction of electronic equipment, H. Baumann, K. Cheng, S. Gut and A. Rossel-Jaeckle for assistance, D. Rigoli for the drawings, T. S. Collett and S. B. Laughlin for reading and improving the manuscript. The work was supported by Swiss NSF grant 3.665-0.84.

Received 26 March; accepted 10 July 1986.

1. von Frisch, K. *Naturwissenschaften* **35**, 38-43 (1948).
2. von Frisch, K. *Experientia* **5**, 142-148 (1949).
3. von Frisch, K. *The Dance Language and Orientation of Bees* (Harvard University Press, Cambridge, Massachusetts, 1967).
4. Laughlin, S. B., Menzel, R. & Snyder, A. W. in *Photoreceptor Optics* (eds Snyder, A. W., Menzel, R. 237-259 (Springer, Berlin, 1975).
5. Wehner, R. & Strasser, S. *Physiol. Ent.* **10**, 337-349 (1985).
6. Helversen, O. & Edrich, W. *J. comp. Physiol.* **94**, 33-47 (1974).
7. Sommer, E. W. Thesis, Univ. Zürich (1979).
8. Labhart, T. *J. comp. Physiol.* **141**, 19-30 (1980).
9. Wehner, R. *Himmelsnavigation bei Insekten. Neurophysiologie und Verhalten, Neujahrsbl. Naturforsch. Ges. Zuerich* **184** (1982).
10. Rossel, S., Wehner, R. & Lindauer, M. *J. comp. Physiol.* **125**, 1-12 (1978).
11. Rossel, S. & Wehner, R. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4451-4455 (1982).
12. Rossel, S., Wehner, R. *J. comp. Physiol.* **A154**, 607-615 (1984).
13. Wehner, R. & Rossel, S. in *Experimental Behavioural Ecology* (eds Hoelldobler, B. & Lindauer, M.) 11-53 (Fischer, Stuttgart, 1985).
14. Waterman, T. H. & Horch, K. W. *Science* **154**, 467-475 (1966).
15. Kirschfeld, K. *Z. Naturforsch.* **27b**, 578-579 (1972).
16. Waterman, T. H. in *Handbook of Sensory Physiology*, Vol. VII/6B (ed. Autrum, H.) 281-469 (Springer, Berlin, 1981).
17. Dyer, F. C. & Gould, J. L. *Science* **214**, 1041-1042 (1981).
18. Waterman, T. H. In *Photoreception and Vision in Invertebrates* (ed. Ali, M. A.) 63-114 (Plenum, New York, 1984).
19. Burghause, F. *Zool. Jb. Physiol.* **83**, 502-525 (1979).
20. Wunderer, H. & Smola, U. *Int. J. Insect Morph. Embryol.* **11**, 25-38 (1982).

Large-scale velocity fields as a test of cosmological models

Nicola Vittorio, Roman Juszkiewicz* & Marc Davis

Department of Astronomy, University of California, Berkeley, California 94720, USA

It has been suggested^{1,4,5-7} that the Hubble expansion on large scales ($\sim 50h^{-1}$ Mpc, where h is the Hubble constant in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$) is distorted by bulk matter flows of surprisingly large amplitude. Using Bayesian statistics of conditional probability, we show here that a high amplitude for large-scale matter flows can be used to rule out, at the 95% confidence level or better, both hot dark matter and cold dark matter models of the growth of structure in the Universe. No random-phase initial conditions with constant-curvature initial perturbations are likely to be consistent with bulk flows relative to the cosmic microwave background (CMB) exceeding 700 km s^{-1} on scales of $50h^{-1}$ Mpc. If the Universe is dominated by cold dark matter and galaxies are a biased mass tracer, then the bulk velocity on this scale should not exceed 150 km s^{-1} .

A decade ago, Rubin *et al.*¹ reported evidence of anisotropy of the Hubble flow on surprisingly large scales. This result has been questioned on several grounds^{2,3}, but re-analyses of the original data^{4,5} and independent observations^{6,7} have yielded similar conclusions. Most recently, Collins *et al.*⁶ have re-observed half of the original Rubin-Ford sample of galaxies using infrared Tully-Fisher (IRTF) techniques, and confirmed the original effect: the Local Group has a velocity of $600 \pm 300 \text{ km s}^{-1}$ relative to a sample of spiral galaxies at a mean redshift of $\sim 5,100 \text{ km s}^{-1}$, in a direction almost orthogonal to the CMB dipole anisotropy. A lower but still large-amplitude bulk motion for a sample of elliptical galaxies at similar distance has been recently found by Burstein *et al.*⁷. However, the Burstein *et al.* motion is directed 44° away from the Collins *et al.* vector after both have been transformed into the co-moving frame; the consistency of the data is therefore open to question. Aaronson *et al.*⁸ have done an extensive study of 10 clusters of galaxies at comparable distance using the IRTF method, and conclude that their mean velocity with respect to the CMB is consistent with zero. However, the clusters that they selected are drawn from the Arecibo observing band, $30^\circ > \delta > 0^\circ$ (where δ is the declination); as the Rubin-Ford vector is nearly orthogonal to this band, this result does not necessarily contradict the others.

The observed strength of galaxy clustering implies that fluctuations on scales larger than $10h^{-1}$ Mpc can be treated as linear perturbations⁹ of a Robertson-Walker metric, if galaxies trace the mass in the Universe or are more tightly clustered than the matter. Large-scale flow fields such as the Rubin-Ford velocity can be accurately treated as linear perturbations. This property is extremely convenient for testing different models for the nature and evolution of the mass distribution in the Universe, because linear perturbation theory is straightforward, tractable and reasonably accurate even for density contrasts of the order of unity. The conventional way of using the observations of the flow field to constrain models is as follows. The initial conditions are set at some very early epoch by choosing the shape and amplitude of the power spectrum of the primordial mass distribution, usually assumed to be a random gaussian field. Then, using the linear theory to propagate the density field in time, one may calculate the resulting root-mean-square (r.m.s.) peculiar velocity on any scale⁹⁻¹².

Here we show that it is possible to tighten the constraints on the existing models by making more detailed comparisons with

the observed departures from a uniform Hubble flow. In particular, it is interesting to compare the expected magnitudes and directions of the velocities, $\mathbf{v}_a$ and $\mathbf{v}_b$, obtained by averaging the flow over two concentric spheres S_a and S_b with radii a and b . Rather than considering spheres with sharp edges, we imagine them to have edges defined by a gaussian smoothing¹³, as defined below. Now consider two measurements of the flow field. The first is the motion of the Local Group relative to the co-moving frame of the Universe as measured by the dipole anisotropy of the CMB¹⁴. A sphere of radius¹⁵ $a \approx 5h^{-1}$ Mpc is moving coherently with the Earth at a velocity of $600 \pm 50 \text{ km s}^{-1}$. The motion of this sphere is strongly influenced by the Virgo cluster, but even here linear perturbation theory specifies the peculiar velocity to reasonable accuracy. On a larger scale we can interpret the Rubin-Ford result as stating that a sphere of radius $b \approx 50h^{-1}$ Mpc is moving at an average velocity given by the vector sum of our microwave dipole velocity and the Rubin-Ford velocity, which is $970 \pm 300 \text{ km s}^{-1}$, directed 40° from the microwave vector. This datum, plus other measurements of streaming velocity on the same scale, are plotted in Fig. 1.

Let us consider the conditional probability $dP(\mathbf{v}_b|\mathbf{v}_a) = p(\mathbf{v}_b, \mathbf{v}_a)d\mathbf{v}_b/p(\mathbf{v}_a)$, where $p(\mathbf{v}_b, \mathbf{v}_a)$ and $p(\mathbf{v}_a)$ are 6- and 3-dimensional joint normal distributions. Our problem is greatly simplified by the statistical isotropy of the flow field. As a consequence, orthogonal spatial components of $\mathbf{v}_b$ and $\mathbf{v}_a$ are uncorrelated. The 6×6 covariance matrix of the distribution $p(\mathbf{v}_b, \mathbf{v}_a)$ is given by:

$$\langle v_a^\alpha v_b^\beta \rangle = \sigma_{ab} \delta_{\alpha\beta} \quad (1)$$

$$\sigma_{ab}^2 = (6\pi^2)^{-1} \Omega^{1.2} H_0^2 \int_0^\infty P(k) \exp\left(-k^2 \frac{(a^2 + b^2)}{2}\right) dk \quad (2)$$

where $\alpha, \beta = 1, 2, 3$ denote different components of velocities and $\delta_{\alpha\beta}$ is the Kronecker symbol. The quantities σ_{ij} correspond to the one-dimensional velocity dispersion, filtered on co-moving scales a, b and $\sqrt{(a^2 + b^2)/2}$ for $i = j = a$, $i = j = b$ and $i \neq j$ respectively. $P(k)$ is the power spectrum of the density perturbations. The relation (2) between the velocity and the density perturbations follows from the linear order continuity equation⁹. The above choice of filtering provides a reasonable fit to the samples of galaxies used in measurements of the peculiar velocity field. The resulting conditional probability is

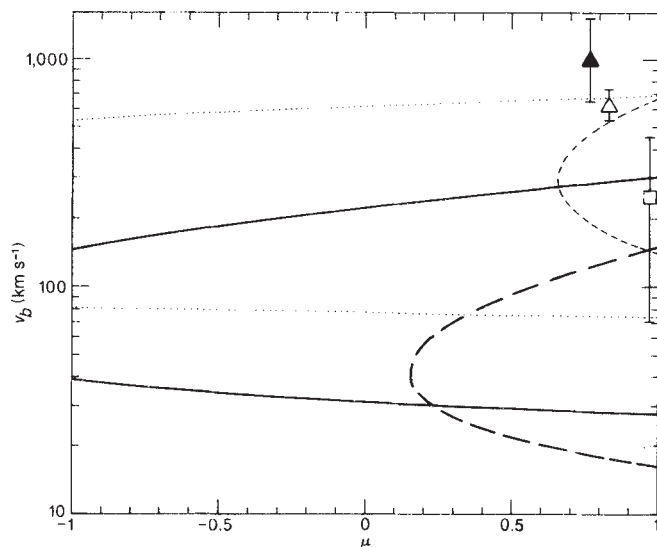
$$dP(\mathbf{v}_b|\mathbf{v}_a) = (2\pi)^{-1/2} \left(\frac{\sigma_{aa}^2}{\Delta} \right)^{3/2} \times \exp \left[\frac{-(\sigma_{ab}^4 v_a^2 + \sigma_{aa}^4 v_b^2 - 2\sigma_{aa}^2 \sigma_{ab}^2 v_a v_b \mu)}{2\Delta \sigma_{aa}^2} \right] v_b^2 dv_b d\mu \quad (3)$$

Here $\mu = (\mathbf{v}_a \cdot \mathbf{v}_b)/(\mathbf{v}_a v_b)$ is the cosine of the misalignment angle and $\Delta = \sigma_{aa}^2 \sigma_{bb}^2 - \sigma_{ab}^4$.

We have considered four models: an $\Omega = 1$ massive-neutrino-dominated universe (HDM), flat and open models dominated by cold dark matter (CDM), in which galaxies trace the mass, and an $\Omega = 1$ CDM model where galaxies are a biased mass tracer¹⁶. (Ω is the density parameter.) In all four cases the initial density fluctuations were assumed to be gaussian, adiabatic and with a Zel'dovich scale-invariant power spectrum ($P(k) \propto k^n$, $n = 1$). We choose a Hubble constant $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ to be consistent with age estimates. Also, such a choice maximizes the theoretical predictions, as higher values of H_0 imply lower peculiar velocities. We evaluated through equation (2) the σ s in different models with the method and normalization procedure described in ref. 18. We compute probability density contours in the (v_b, μ) plane from equation (1) and show in Fig. 1 the contours containing 95% of the integrated probability density. None of the models is consistent with the Rubin-Ford data¹. The biased CDM model is strongly inconsistent with both

* Permanent address: Warsaw University Observatory and Copernicus Astronomical Center, Poland.

Fig. 1 The 95% confidence contour in the (v_b, μ) plane for different cosmological models. Here, v_b is the magnitude of the bulk motion of a sphere of radius $50h^{-1}$ Mpc with respect to the CMB; μ is the cosine of the misalignment angle between $\mathbf{v}_b$ and the apex of the CMB dipole anisotropy. The dotted line applies to an $\Omega = 1$ HDM model, the solid line to an $\Omega = 1$ CDM model, the short-dashed line to an $\Omega = 0.1$ CDM model, and the long-dashed line to an $\Omega = 1$ CDM model with bias. $\blacktriangle$, from ref. 6; $\triangle$, from ref. 7; $\square$, from ref. 8). The error bars in each case are the quoted 1σ errors; formal errors have not yet been quoted for the data from ref. 7 and in this case the error bar is only approximate.



the Rubin-Ford data and the Burstein *et al.*⁷ result, but is consistent with the Aaronson *et al.*⁸ data.

In the $\Omega = 1$ HDM model ($\sigma_{aa} = 1,040 \text{ km s}^{-1}$; $\sigma_{bb} = 230 \text{ km s}^{-1}$) the conditional probability of measuring $v_b = 970 \text{ km s}^{-1}$ given $v_a = 600 \text{ km s}^{-1}$ turns out to be unacceptably small. The observational point of Collins *et al.*⁶ lies outside the 95% confidence level in Fig. 1. The observations of Burstein *et al.*⁷ are marginally consistent with the prediction at the same confidence level. However, if we accept the HDM model, then our Local Group occupies an unusually quiet spot in the Universe. The probability of measuring $0 \leq v_a \leq 600 \text{ km s}^{-1}$ on a scale $a = 5h^{-1}$ Mpc is only 5%. Note also the very weak alignment between $\mathbf{v}_a$ and $\mathbf{v}_b$ in Fig. 1.

For the $\Omega = 1$ CDM universe ($\sigma_{aa} = 276 \text{ km s}^{-1}$; $\sigma_{bb} = 90 \text{ km s}^{-1}$), the conditional probability of measuring $v_b = 970 \pm 300 \text{ km s}^{-1}$ given $v_a = 600 \text{ km s}^{-1}$ is even smaller than in the HDM case. Moreover, the r.m.s. velocity of $50h^{-1}$ Mpc falls far short of the observed value.

The biased CDM model^{13,16} has been the most successful to date in reconciling the inflationary prejudices for a flat universe with the observations. However, it has extreme difficulties in generating large-scale high peculiar velocity^{12,17}. For this model σ_{aa} and σ_{bb} are expected to be roughly a factor of 3 smaller than their values in the $\Omega = 1$ CDM model without bias. Figure 1 shows that in a biased galaxy formation model, v_b is hopelessly small. Moreover, the alignment between $\mathbf{v}_a$ and $\mathbf{v}_b$ is stronger than expected in the unbiased CDM model. At the 95% confidence level, we can exclude values of $v_b > 150 \text{ km s}^{-1}$ and misalignments $> 80^\circ$. Thus, the biased CDM model is consistent only with the Aaronson *et al.*⁸ data.

The open CDM model appears to be the most favourable candidate for explaining high peculiar velocities on a large scale. In this case, the amplitude of the density fluctuations (and hence, peculiar velocities) on large scales increases¹⁸ with decreasing Ω . Unfortunately, even in this case the 95% confidence level

excludes $v_b > 560 \text{ km s}^{-1}$ for $\Omega = 0.1$, the lower bound on the mass density. Note that the 95% confidence contour in Fig. 1 contains a narrow range of μ . This indicates a higher chance of alignment between $\mathbf{v}_a$ and $\mathbf{v}_b$, which is a consequence of the very large coherence length in the matter distribution. The same condensations responsible for the peculiar velocity on large scales are largely responsible for the peculiar velocity on small scales. Therefore, $\mathbf{v}_a$ and $\mathbf{v}_b$ tend to be naturally aligned.

The existing observations of v_b have large error bars and low significance, as seen in Fig. 1. It is apparent that a large amplitude for the bulk velocity on large scales is inconsistent with all random-phase models having scale-invariant initial perturbations. Alternative models, such as isothermal initial conditions with a spectral slope of $n = -1$, or antibiased models, could be fabricated to meet this restriction, but such models are highly artificial and unable to meet other observational constraints such as the microwave background fluctuation limits. Confirmation of a high amplitude for v_b would imply either that (1) inflation in its usual form¹⁹ is wrong because it leads to scale-invariant, random-phase initial conditions; (2) that non-random-phase fluctuations²⁰ were somehow specified in the early Universe (cosmic strings^{21,22} are the only mechanism currently discussed for generation of non-gaussian fluctuations, but it is not yet clear how to test this model against observations); or that (3) in spite of confirmation by several observers using different techniques, the Rubin-Ford effect may not be a measurement of the velocity of a shell of galaxies. The signal may instead be the result of bias induced by substantial spatial inhomogeneity in the Rubin-Ford sample^{2,23}. We can conclude from our analysis that further attempts to improve the significance and consistency of the observations are justified because they offer a powerful cosmological probe.

This research was supported in part by NSF grant AST-8642105. N.V. acknowledges financial support from the European Space Agency and DOE grant 84-ER40161.

Received 17 April; accepted 26 June 1986.

- Rubin, V. C., Ford, W. K., Thonnard, N., Roberts, M. S. & Graham, J. A. *Astr. J.* **81**, 687-737 (1976).
- Fall, S. M. & Jones, B. J. T. *Nature* **262**, 457-470 (1976).
- Tammann, G., Yahil, A. & Sandage, A. *Astrophys. J.* **234**, 775-784 (1979).
- Schechter, P. L. *Astr. J.* **82**, 569-576 (1977).
- Peterson, C. J. & Baumgart, C. W. *Astr. J.* **91**, 530-538 (1986).
- Collins, A., Joseph, R. D. & Robertson, N. A. *Nature* **320**, 506-508 (1986).
- Burstein, D. *et al. Galaxy Distances and Deviations from Universal Expansion* (eds Madore, B. F. & Tully, R. B.) (NATO, Brussels, 1986).
- Aaronson, M. *et al. Astrophys. J.* **302**, 536-563 (1986).
- Peebles, P. J. E. *The Large Scale Structure of the Universe* (Princeton University Press, 1980).
- Clutton-Brock, M. & Peebles, P. J. E. *Astr. J.* **86**, 1115-1119 (1981).
- Kaiser, B. *Astrophys. J.* **273**, L17-L20 (1983).

- Vittorio, N. & Silk, J. *Astrophys. J.* **293**, L1-L4 (1985).
- Bardeen, J. M., Bond, J. R., Kaiser, N. & Szalay, A. S. *Astrophys. J.* **304**, 15-61 (1986).
- Sandage, A. *Astrophys. J.* (in the press).
- Lubin, P. & Vilella, T. *Galaxy Distances and Deviations from Universal Expansion* (eds Madore, B. F. & Tully, R. B.) (NATO, Brussels, 1986).
- Davis, M., Efstathiou, G., Frenk, C. & White, S. *Astrophys. J.* **292**, 371-394 (1985).
- Bond, J. R. *Galaxy Distances and Deviations from Universal Expansion* (eds Madore, B. F. & Tully, R. B.) (NATO, Brussels, 1986).
- Vittorio, N. & Turner, M. Preprint Univ California, Berkeley (1986).
- Brandenberger, R. H. *Rev. mod. Phys.* **57**, 1-60 (1985).
- Peebles, P. J. E. *Astrophys. J.* **274**, 1-6 (1983).
- Zel'dovich, Ya. B. *Mon. Not. R. astr. Soc.* **192**, 663-666 (1980).
- Vilenkin, A. *Phys. Rep.* **121**, 263-315 (1985).
- Chincarini, G. & Rood, H. *Astrophys. J.* **230**, 648-654 (1979).

A new infrared spectral component of the quasar 3C273

E. I. Robson*, W. K. Gear*, L. M. J. Brown*,
T. J.-L. Courvoisier†‡, M. G. Smith‡, M. J. Griffin§
& A. Blecha||

* School of Physics & Astronomy, Lancashire Polytechnic,
Preston PR1 2TQ, UK

† ST-ECF, European Space Observatory, Karl-Schwarzschild-Str. 2,
D-8046 Garching bei München, FRG

‡ UK Infrared Telescope, 665 Komohana Street, Hilo,
Hawaii 96720, USA

§ Department of Physics, Queen Mary College, Mile End Road,
London E1 4NS, UK

|| Geneva Observatory, CH 1290 Sauverny, Geneva, Switzerland

Following the dramatic infrared to millimetre-wavelength flare seen in the quasar 3C273 during 1983¹, we have continued to monitor its overall continuum emission. Recent measurements show that the 10- μ m to 3-mm emission has decayed to a level well below any seen previously^{2,3}, while the 1-4- μ m emission has remained relatively constant. This behaviour has revealed the presence of an apparently non-variable component which dominates the near-infrared emission in 3C273 and includes the small 'bump' at $\sim 3.5\mu$ m in the power-law continuum previously noted by Neugebauer *et al.*³. The origin of this component is probably not thermal re-radiation by dust grains but may be due to free-free emission from very dense, broad-line clouds⁴.

In 1982, 3C273 became the first quasar to have its infrared to radio spectrum fully determined⁵. At that time there had been no evidence of any significant variability of the 1,100- μ m (ref. 2) or 2.2- μ m (ref. 3) fluxes over a period of several years. This spectrum is shown in Fig. 1a, and shall henceforth be referred to as the 'quiescent' spectrum. Since that time we have continued to monitor 3C273 at all wavelengths (T.J.-L.C. *et al.*, in preparation).

In 1983, 3C273 underwent a dramatic flare in its infrared to millimetre emission¹. A spectrum taken during the flare in March 1983 is shown in Fig. 1b. Marscher and Gear⁶ have shown that the behaviour of this flare was inconsistent with all previous models, but that a new model based on shock waves travelling in an adiabatic relativistic jet could reproduce the data very well.

Observations of 3C273 are possible only during certain times of the year, the observing season extending from December to July. Our observations in early 1985 showed that the infrared and millimetre emission had returned to the 'quiescent' level. Subsequent near-infrared measurements (J, H, K and L bands) throughout the 1985 observing season showed that the infrared fluxes remained constant; however, no millimetre to sub-millimetre 10-20- μ m observations were obtained. At the beginning of the 1986 observing season, in January, we found that the millimetre and sub-millimetre fluxes were lower by about a factor of three than the quiescent level, while the JHK fluxes had remained constant within the uncertainties. In February 1986, we obtained 10- and 20- μ m data, as well as 1,100- and 800- μ m and JHKL measurements. These showed that the 10- μ m flux was a factor of ~ 2 below the quiescent level. The February 1986 spectrum is shown in Fig. 1c. A further set of 1- μ m to 1-mm data taken on March 5 shows a possible decline in the 20- μ m flux, with no significant change in the 800-1,100- μ m or JHKL fluxes since the February data. The March spectrum is shown in Fig. 1d, which also contains continuum measurements at 1.3, 2.0 and 3.3 mm from the NRAO (National Radio Astronomy Observatory) 12-m telescope at Kitt Peak between 2 and 6 March. A further measurement at 10 μ m on 31 March

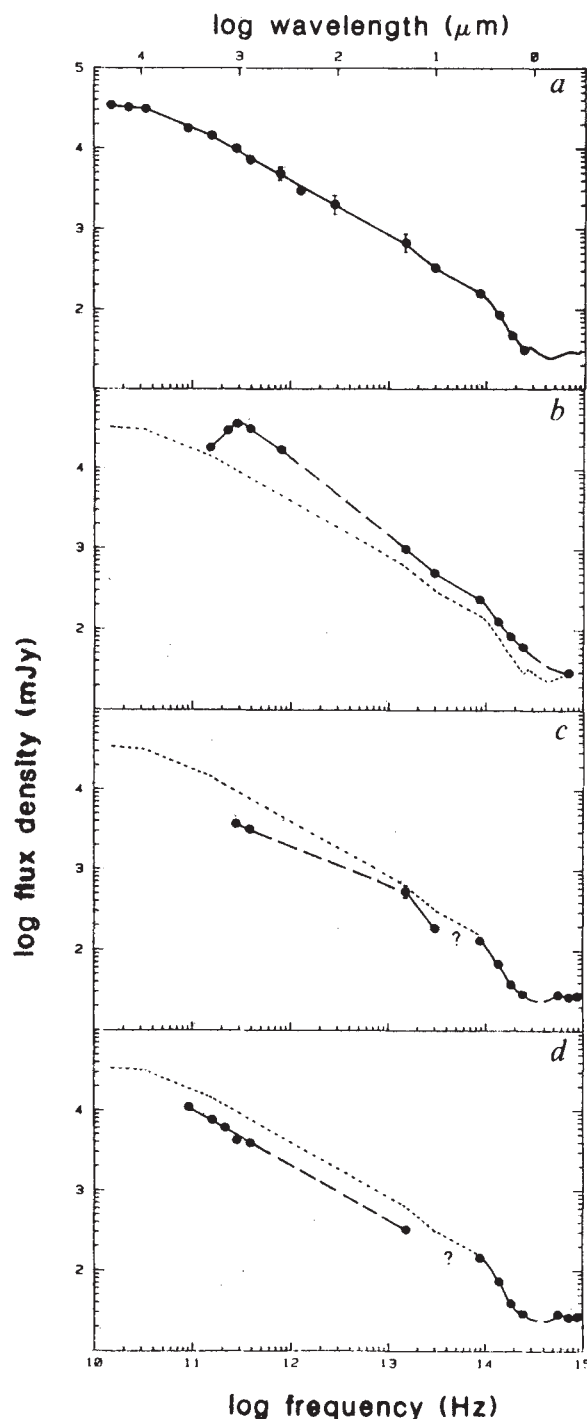


Fig. 1 History of the millimetre to optical spectrum of 3C273: a, the 'quiescent' spectrum⁵; b, 4-7 March 1983, flare spectrum¹; c, 15-24 February 1986 (this work); d, 3-6 March 1986 (this work). The quiescent spectrum is shown for comparison as a dashed line in b-d. The differences between the 0.36-3.45- μ m fluxes in c and d and the quiescent values are too small to be seen on this diagram.

gave a value of 197 ± 20 mJy. UVB optical Geneva photometry was also performed during December-March. These results showed that the V-band flux remained constant, while the U-band flux increased by 15% between December and February (see Table 1). The infrared to millimetre observations were all made at the United Kingdom Infrared Telescope (UKIRT), using common-user instrumentation; the optical observations were made at La Silla, using the 70-cm Swiss telescope.

An obvious feature of the February 1986 spectrum in Fig. 1c is that the JHKL emission lies above an extrapolation of the

Table 1 Flux densities for 3C273 (mJy)

Wavelength (μm)	0.36	0.43	0.55	1.25	1.65	2.20	3.45	10.5	20	770	1,070	1,320	2,000	3,350
Frequency (THz)	845	703	548	240	182	131	86	29	15	0.39	0.28	0.23	0.15	0.09
19 Dec. 1985	23.8 (0.3)	24.9 (0.3)	28.0 (0.3)	—	—	—	—	—	—	—	—	—	—	—
6 Jan. 1986	—	—	—	28.7 (2.3)	39.2 (3.1)	73 (6)	—	—	—	3,100 (500)	3,700 (400)	—	—	—
9 Jan. 1986	26.9 (0.3)	26.9 (0.3)	28.7 (0.3)	—	—	—	—	—	—	—	—	—	—	—
15 Feb. 1986	27.2 (0.3)	26.6 (0.3)	28.5 (0.3)	—	—	—	—	—	—	—	—	—	—	—
19 Feb. 1986	—	—	—	28.2 (0.8)	37.6 (1.1)	66 (2)	131 (4)	—	—	—	—	—	—	—
20–21 Feb. 1986	—	—	—	—	—	—	—	—	—	3,250 (350)	3,800 (400)	—	—	—
24 Feb. 1986	—	—	—	28.6 (1.0)	39.3 (1.3)	73 (3)	139 (5)	187 (23)	540 (100)	—	—	—	—	—
3–5 Mar. 1986	—	—	—	—	—	—	—	—	—	—	—	—	7,600 (800)	10,800 (1,000)
6 Mar. 1986	26.7 (0.3)	26.4 (0.3)	28.9 (0.3)	29.8 (1.5)	40.6 (2.1)	75 (4)	140 (7)	—	325 (30)	3,900 (400)	4,200 (400)	6,000 (600)	—	—

Numbers in parentheses are $\pm 1\sigma$ errors.

20- to 10- μm slope. This was marginally indicated in the quiescent spectrum of Fig. 1a, and is now confirmed by the present measurements. We must distinguish between two differing interpretations of the overall millimetre to optical continuum spectrum. Either there is a single, featureless emission spectrum which becomes absorbed in the mid-infrared region, or there are two distinct processes—one dominating the 3-mm to 10- μm emission, and the other the 1–4- μm emission.

One obvious way of suppressing mid-infrared emission from an otherwise featureless spectrum is to invoke absorption due to silicates. If extensive dust were being formed in the outer regions of the quasar nucleus, then the 1–4- μm emission would be expected to decrease in concert with the 1,100–20- μm emission, and also simultaneously to steepen due to reddening by the silicate grains. In fact, the 1–4- μm emission remains essentially constant. The optical spectrum should show an even greater effect of reddening, but again there is no indication of any spectral steepening; in fact, the opposite is observed. No evidence has been found for dust features commonly seen at 3.5 μm (ref. 7). Free-free or synchrotron self-absorption are also possible mechanisms for producing spectral cutoffs at longer wavelengths. However, in both cases, once a high-frequency spectrum has become absorbed it remains absorbed to all lower frequencies, and one then violates the initial supposition that the millimetre to infrared emission is a single spectral component.

We therefore consider it more likely that there are indeed two separate components involved. We note that the February 1986 20- μm flux is close to the quiescent level; however, as we have no 20- μm measurements from 1984 to 1986, we do not know what its behaviour was in the interim. The simplest explanation of the radio to mid-infrared continuum is a decaying synchrotron spectrum (refs 5, 8 and T.J.-L.C. *et al.*, in preparation). The 10- and 20- μm fluxes are the high-frequency tail of this spectrum which, in the absence of an injection of new particles, will show a cutoff due to radiative losses. In the context of a relativistic jet model, the change in the overall flux level from the quiescent case to what we see now may reflect a decrease in the injection rate of relativistic particles, perhaps due to a lull in the activity of the central engine. The associated decrease in the cutoff frequency would then be explained by a decrease in the maximum energy of the radiating electrons—note that an increase in magnetic field could also produce this effect but would lead to an increase in the self-absorption turnover frequency, which is not observed.

An implication of a two-component model is that the varying

radio to mid-infrared component must steepen considerably between 10 and 4 μm in all but the flare spectrum, in order for it to contribute little observed flux at 1–4 μm and shorter wavelengths. This may affect models of the optical to ultraviolet emission that depend on subtracting an underlying power-law continuum (see ref. 8). The presence of a 'bump' at 3.5 μm in the spectrum of 3C273 has already been noted^{3,9–11}; now that the millimetre to infrared synchrotron component has decreased below the quiescent level, this bump is more clearly revealed.

What is the origin of this apparently non-variable near-infrared component, which must peak not much longward of 4 μm ? First, given that the millimetre to mid-infrared spectrum is suggested to be synchrotron emission, we consider the possibility that the 1–4- μm emission is a separate component, but also of synchrotron origin. Synchrotron emission can certainly produce the observed power-law spectral shape; the turnover near 4 μm could be due to self-absorption. However if this were the case, the emitting region would have to be extremely compact, of the order of a few astronomical units, assuming a maximum brightness temperature close to the Compton limit and Hubble constant $H_0 = 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$. The resulting derived magnetic field strength is $B \approx 100 \text{ G}$. One would then expect to see rapid variability at these wavelengths, quite the opposite of the observed situation³ (although we note that as yet there has been no firm upper limit set on very rapid, that is, hourly or less, variability in the infrared emission from 3C273). If the turnover were due to free-free absorption in the broad-line region, then the synchrotron-emitting region need not be so compact; however, as has been pointed out before¹², this would imply a covering factor for the broad-line clouds of unity, whereas the recombination-line observations suggest a very much lower value. A low-energy cutoff in the electron energy distribution will produce a spectral cutoff with a low-frequency spectral slope of 1/3. This cannot be ruled out by our data. We note, however, that synchrotron radiation would be expected to produce substantial polarization, as is seen in blazar-type objects, whereas the infrared polarization in 3C273 is very low ($<1\%$; ref. 13). This low polarization compared with the measured 3.3-mm polarization of 5.8% (ref. 14) also argues against an origin due to synchrotron self-Compton scattering, as does the lack of simultaneous variability of the two spectral regions.

Second, we consider thermal re-radiation from heated dust grains. Using an appropriate radial distribution of grain number density, it would in principle be possible to reproduce the observed approximate infrared power-law behaviour. However,

in order to radiate significant energy at wavelengths shorter than 2 μm , the grain temperature would have to exceed 2,000 K, well in excess of the maximum temperature permitted by current models of interstellar grains. Even if grains can exist at these temperatures, there is a further problem in that the dust should also manifest itself by reddening the observed recombination lines. All evidence suggests that there is little or no reddening of the recombination lines in 3C273 (see, for example, ref. 15). Furthermore, as mentioned above, no confirmed evidence of a 3.5- μm dust feature has been found in 3C273⁷.

Thermal comptonization has previously been suggested as a contributor to the infrared continuum of both radio-loud and radio-quiet quasars⁹. Comptonization of a source of soft photons by the broad-line-region gas can produce an approximate power-law (see Fig. 6 of ref. 10). However, this merely raises the question of where the soft photons originated from. There has never been observational evidence of a source of far-infrared seed photons to produce thermal Compton scattering in the observed spectrum⁵, although we have no observations of the 30–300- μm spectrum at the current epoch.

One further possibility is that the infrared spectrum does indeed arise in the broad-line-region but is free-free emission from very dense clouds ($n \approx 10^{11} \text{ cm}^{-3}$). This suggestion was originally put forward by Puetter and Hubbard⁴ to explain the very small bump seen in 3C273 in its base state, and also the low Paschen- α fluxes in some objects¹⁵. Their curves are at least qualitatively consistent with the shape of the bump. This process needs to be investigated further.

We believe that we have seen in these observations of 3C273 the decay of a 'normal' synchrotron radio spectrum. This may well be an aftermath of the flare (although perfect temporal coverage was not possible), and may imply that 3C273 is now 'recharging its batteries'. If the timescale of recharging is sufficiently long, the core of 3C273 could evolve from a radio-loud to a radio-quiet quasar.

If the ionizing continuum in 3C273 were to switch off, we might expect to see a change in the line emission after several years. This situation would then be analogous to the change seen in NGC4151, whose optical spectrum has changed from that of a Seyfert I to that of a Seyfert II¹⁷. It will be extremely important to continue to monitor 3C273 at all wavelengths. We may also expect to see a recurrence of the flare phenomenon.

In spite of an intensive monitoring campaign, 3C273 continues to be enigmatic. Our temporal coverage is still insufficient to yield a complete understanding of this object's behaviour. Our two-component model requires further observational investigation, in particular, more detailed measurement of the shape of the spectrum, especially between 20 and 4 μm .

We thank the UKIRT, NRAO and Geneva Observatory staff for assistance with the observations. T.J.-L.C. acknowledges useful discussions with M.-H. Ulrich, W.K.G., M.J.G. and L.M.J.B. acknowledge financial support from SERC. The National Radio Astronomy Observatories is operated by Associated Universities Inc., under contract from the NSF.

Received 28 April; accepted 26 June 1986.

1. Robson, E. I. *et al.* *Nature* **305**, 194–196 (1983).
2. Gear, W. K. *et al.* *Astrophys. J.* **280**, 102–106 (1984).
3. Neugebauer, G. *et al.* *Astrophys. J.* **230**, 79–84 (1979).
4. Puetter, R. C. & Hubbard, E. N. *Astrophys. J.* **295**, 394–401 (1985).
5. Clegg, P. E. *et al.* *Astrophys. J.* **273**, 58–63 (1983).
6. Marscher, A. P. & Gear, W. K. *Astrophys. J.* **298**, 114–127 (1985).
7. Lee, T. J. *et al.* *Nature* **295**, 214–215 (1982).
8. Camenzind, M. & Courvoisier, T. J.-L. *Astr. Astrophys.* **140**, 341–344 (1984).
9. Ennis, D. J., Neugebauer, G. & Werner, M. W. *Astrophys. J.* **262**, 460–477 (1982).
10. Malkan, M. A. *Astrophys. J.* **268**, 582–590 (1983).
11. Cutri, R. M., Wiziniewski, W. Z., Rieke, G. H. & Lebofsky, M. J. *Astrophys. J.* **296**, 423–429 (1985).
12. Condon, J. J., O'Dell, S. L., Puschell, J. J. & Stein, W. A. *Astrophys. J.* **246**, 624–646 (1981).
13. Stockman, H. S., Moore, R. L. & Angel, J. R. P. *Astrophys. J.* **279**, 485–498 (1984).
14. Rudnick, L. *et al.* *Astrophys. J.* **225**, L5–L9 (1978).
15. Davidson, K. & Netzer, H. *Rev. mod. Phys.* **51**, 715–866 (1979).
16. Lacy, J. H. *et al.* *Astrophys. J.* **256**, 75–82 (1982).
17. Penston, M. V. & Perez, E. *Mon. Not. R. astr. Soc.* **211**, 33P–39P (1985).

The flux of meteoroids and orbital space debris striking satellites in low Earth orbit

M. R. Laurance & D. E. Brownlee

Department of Astronomy, University of Washington, Seattle, Washington 98195, USA

Analysis of the large number of high-velocity impact craters found on surfaces of the Solar Maximum Mission spacecraft is providing an unprecedented opportunity to measure the near-Earth meteoroid flux over a wide range of particle sizes. Zook¹ and Schramm *et al.*² have reported size frequencies and compositions for particles that produced 40–520- μm penetration holes in the multi-layer plastic thermal insulation blanket from 'Solar Max'. Here we report the results of analysis of 331 craters in the 0.09–250 μm size range, found in the solid aluminium of the thermal control louvers. By elemental analysis of particle residue in the craters and scanning electron microscope imaging, we were able to distinguish impacts by natural meteoroids from those of orbital spacecraft debris, and directly determine the time-averaged flux and size distribution of both particle types over the mass range 10^{-13} – 10^{-7} g. The most common natural meteoroid craters contain residue that has chondritic elemental composition, while the debris craters contain a variety of residue types including paint. For 10^{-12} g particles the debris flux exceeds the natural meteoroid flux by two orders of magnitude, but at 10^{-8} g the fluxes are similar.

On 10 April 1984 the crew of Space Shuttle flight 41-C recovered surfaces from the Solar Maximum Mission spacecraft. The surfaces had been exposed to meteoroid bombardment for 4.15 yr, at orbital altitudes ranging from 500 to 570 km. Among the materials recovered from Solar Max were thermal control louvers, measuring 4.5×25 cm, made of polished 99%-purity aluminum sheet ~ 130 μm thick. The major impurity is iron, occurring as sub-micrometre-sized inclusions. The surface quality of the four louvers that we studied is excellent, and the scarcity of crater-like artefacts permits the straightforward location of craters as small as 0.1 μm in diameter. Craters were

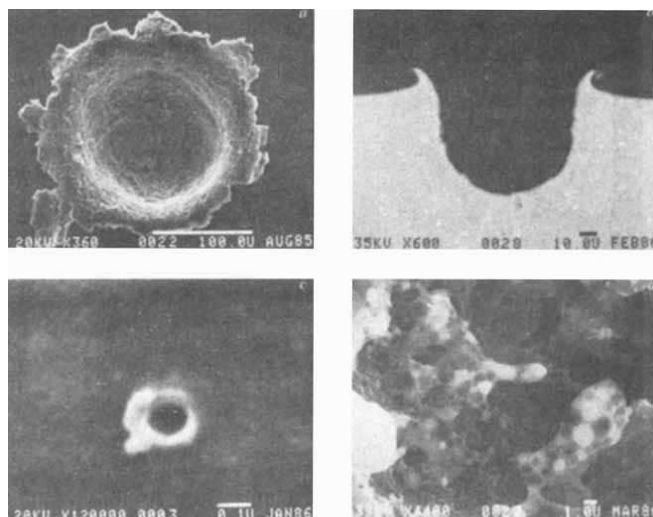


Fig. 1 Scanning electron micrographs of high-velocity craters. *a*, 150- μm -diameter crater containing chondritic residue (scale bar 100 μm). *b*, Polished cross-section of a 90- μm crater with chondritic residue (scale bar 10 μm). *c*, 0.09- μm crater (scale bar 0.1 μm). *d*, Frothy residue in the bottom of a 205- μm crater (scale bar 1 μm). The residue consists of bright blebs of sulphide and Fe–Ni metal, enclosed in a mixture of chondritic material and aluminium.

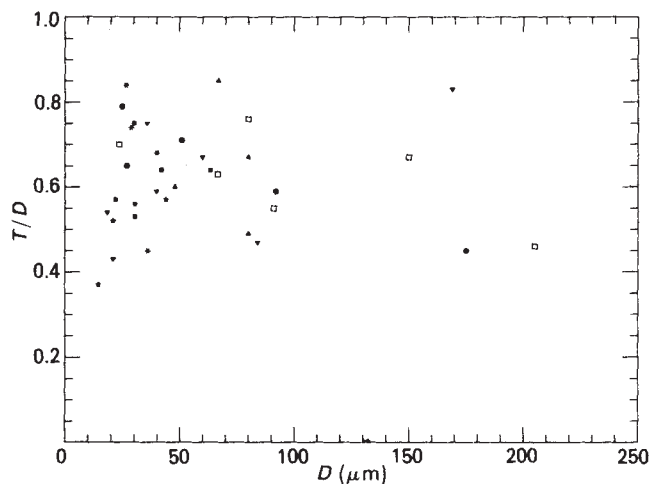


Fig. 2 Depth/diameter ratios (T/D) determined by optical scanning microscopy. The crater diameter D is defined at the line of intersection of the crater and the original surface plane. Symbols for extraterrestrial craters: Δ , sulphide; $\bullet$, mafic; $\square$, chondritic craters. Non-extraterrestrial craters: $\blacksquare$, paint; ∇ , Fe, Si; $\star$, miscellaneous craters; $\ast$, craters with no residue.

located using an optical stereo microscope at magnifications up to $\times 200$, and a scanning electron microscope (SEM) at three discrete search magnifications: $\times 100$, $\times 200$ and $\times 1,000$. For each of the four searches, the number of craters found, the area scanned and the minimum crater diameter that could be reliably detected are as follows: optical, 65 craters, 438 cm^2 , $20 \mu\text{m}$; SEM $\times 100$, 151 craters, 100 cm^2 , $5 \mu\text{m}$; SEM $\times 200$, 11 craters, 1.57 cm^2 , $3 \mu\text{m}$; SEM $\times 1,000$, 90 craters, 0.31 cm^2 , $0.5 \mu\text{m}$. All of the craters are bowl-shaped, with the raised, petalled rims (Fig. 1) which characterize high-velocity impacts into metal. The distribution of depth/diameter ratios (Fig. 2) is peaked, and the general lack of shallow craters indicates that the crater population is probably not 'contaminated' by low-velocity impacts. The depth/diameter ratios are similar to those found for silicate lunar samples and surfaces cratered in laboratory simulations³. The measured crater densities plotted in Fig. 3 are consistent with published data for penetration holes in the Solar Max thermal blanket¹.

All but 13.6% of the craters contained particle residue detectable in the SEM with a standard solid-state (Si-Li) X-ray detector. The craters without residue may contain Al, or elements lighter than Na which would not be detectable with our X-ray system. The most common crater liners are composed of various concentrations of Ti, Zn, Cl, Si and K, or Ti, Si and Fe. Based on the analysis of impacts in the thermal blanket¹, we assume that these craters were made by high-velocity debris from other spacecraft, most probably paint. Twenty of the craters contain residue whose composition is consistent with impacts of extraterrestrial material. We group these compositions into chondritic, sulphide and mafic, the three most common classes of micro-meteorites in the stratosphere⁴. The relative elemental composition of the chondritic craters matches solar abundance for Mg, Si, S, Ca, Cr, Mn, Fe and Ni within a factor of three. The sulphides are dominated by an iron-nickel sulphide composition, with some chondritic fractions, and the mafic craters (MSF of ref. 2) contain Mg, Si and Fe, but no S or Ni. We assume that the mafic craters are extraterrestrial, but their composition is certainly not unique to meteoritic materials. The morphology of the crater residue ranges from vesicular mixtures of glass, sulphides and metal to essentially smooth films, and is similar to that seen in some experimental impact craters⁵. Optically, most of the extraterrestrial residue is black. Quantitative analysis of the residues in extraterrestrial craters, combined with crater-

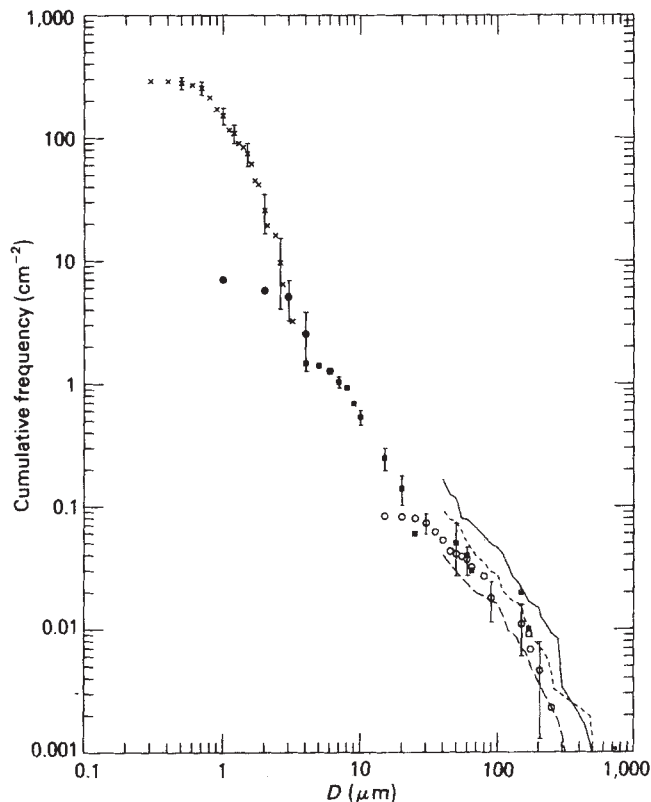


Fig. 3 Measured crater frequencies from the Solar Max louvers (points), compared with data from penetrations in the multi-layer thermal insulation blanket¹ (lines). SEM scans: $\times$, $\times 1,000$; $\bullet$, $\times 200$; $\blacksquare$, $\times 100$; $\circ$, Optical scan. —, Blanket section 9; ---, blanket section 6; — · —, main electronics box blanket.

ing theory, indicates that the fraction of surviving meteoroids in the craters ranges from 3 to 48%.

Figure 4 shows fluxes for natural meteoroids and spacecraft debris. Although there was some geometric shielding for the louvers, the effect was small ($<25\%$) and was not included in the computation. There is no foreshortening effect of the louvers, as they were oriented parallel to the spacecraft exterior (W. Ousley, personal communication) during exposure. Masses of crater-forming projectiles were computed with the penetration depth formula developed in ref. 6:

$$T = \frac{1}{(\epsilon^{0.06} \rho_T^{0.5})} m^{0.4} (v \cos \alpha)^{0.88} \rho_P^{0.33} \quad (1)$$

where T , ϵ , ρ_T , v , α and ρ_P are the crater depth (cm), the target ductility, the target density (g cm^{-3}), the projectile mass (g), the projectile velocity (km s^{-1}), the angle from normal of impact and the projectile density (g cm^{-3}). Using our observed mean crater depth/diameter ratio of 0.62, and the value of $\epsilon^{0.06} \rho_T^{0.5} = 1.49$ for Al (ref. 6), equation (1) can be rearranged to relate the projectile mass to crater diameter, D :

$$m = \left[\frac{0.62 D}{0.67 (v \cos \alpha)^{0.88} \rho_P^{0.33}} \right]^{2.5} \quad (2)$$

To construct Fig. 4 we chose $\alpha = 45^\circ$ and $\rho_P = 2.5 \text{ g cm}^{-3}$ for all particles, and $v = 20 \text{ km s}^{-1}$ (ref. 7) for the extraterrestrial impacts and $v = 8 \text{ km s}^{-1}$ for space debris. All of the comparative data from other sources have been modified to apply to what would be measured by a flat, spinning plate in low Earth orbit, essentially the exposure conditions for Solar Max. Where appropriate, the data were corrected for Earth shielding and gravitational focusing. The gravitational focusing factor used was 1.46, corresponding to an initial velocity of 16.57 km s^{-1} relative to the Earth.

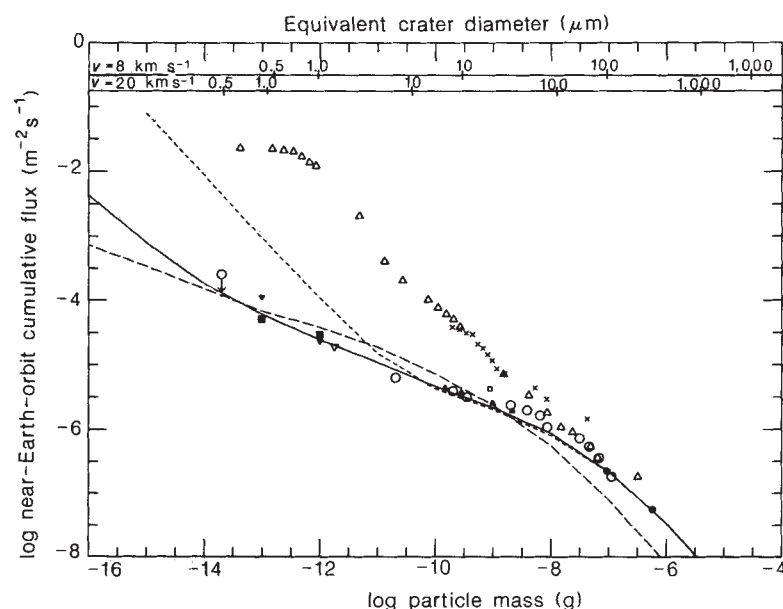


Fig. 4 Flux of extraterrestrial particles and spacecraft debris in low Earth orbit. All of the plotted space measurements have been adjusted to be appropriate for what would be measured by a tumbling flat-plate collector in low Earth orbit, the exposure conditions for the Solar Maximum Mission spacecraft. Mass computations assume spherical particles of density 2.5 g cm^{-3} , with impact velocities (v) of 20 km s^{-1} for extraterrestrial meteoroids and 8 km s^{-1} for spacecraft debris. $\circ$, Solar Max extraterrestrial flux; $\triangle$, Solar Max debris flux; $\times$, Skylab IV/Apollo windows⁹; $\blacktriangle$, Explorer XVI, XXIII¹¹; $\bullet$, Pegasus¹¹; $\square$, Skylab IV S228 experiment¹²; ∇ , Pioneer 8, 9¹³; $\blacksquare$, HEOS II¹⁴; ∇ , micro-abrasion foil experiment¹⁰; —, —, interplanetary flux from refs 7 and 15, respectively; - - -, lunar flux from ref. 7.

Our measured extraterrestrial flux is generally consistent with the spacecraft detector data and model fluxes plotted in Fig. 4. The significant deviation from lunar microcrater data for small craters supports the suggestion⁸ that the lunar microcrater record overestimates the flux of small meteoroids because of high-velocity secondary ejecta derived from impacts on the lunar surface. Our orbital debris flux is consistent with Skylab window data⁹ and with the data from the Solar Max thermal blanket¹. These results indicate that the omnidirectional spacecraft debris and meteoroid fluxes are similar for microgramme particles, but that debris strongly dominates in the picogramme range. Although we cannot rule out detection problems, the fall-off of the flux curve below 1 pg appears to be real. Our fluxes of small particles are much higher than those measured by the micro-abrasion foil experiment¹⁰ on the STS-3 shuttle flight. This experiment was exposed for 8 days at an altitude of 241 km , and pointed at the zenith for most of the exposure. The discrepancy between the two measurements could be an indication that the omnidirectional debris flux is either variable in time or is a strong function of altitude. At lower altitudes the debris flux should diminish due to loss on re-entry.

The data presented here show that craters in appropriate metal targets can be used to determine the fluxes and compositions for a wide range of particle sizes. This technique will be used on Earth-orbiting meteoroid experiments such as the Long Duration Exposure Facility (LDEF), and could be used on a comet-sample return mission involving a fast coma fly-by followed by Earth return of collected material.

We thank W. Ousley of the Goddard Spaceflight Center for loaning us the louvers and P. D. Nelson and the crew of shuttle flight 41-C for recovering them from space. We also thank F. Horz, H. Zook, D. S. McKay and P. Tsou for many helpful suggestions. This work was supported by NASA-JSC grant NAG 9-124 and JPL contract 956257.

Received 7 April; accepted 16 June 1986.

1. Zook, H. A. *Proc. SMRM Degradation Study Workshop*, 247-263 (NASA, Greenbelt, 1985).
2. Schramm, L. S., McKay, D. S., Zook, H. A. & Robinson, G. A. *Proc. lunar planet. Sci. Conf.* **16**, 434-435 (1985).
3. Fechtig, H., Grün, E. & Kissel, J. in *Cosmic Dust* (ed. McDonnell, J. A. M.) 607-664 (Wiley, Chichester, 1978).
4. Brownlee, D. E. *A. Rev. Earth planet. Sci.* **13**, 147-173 (1985).
5. Horz, F., Fechtig, H., Janicke, J. & Schneider, E. *Proc. lunar planet. Sci. Conf.* **14**, B353-B363 (1983).
6. Pailer, N. & Grün, E. *Planet. Space Sci.* **28**, 321-331 (1980).
7. Grün, E., Zook, H. A., Fechtig, H. & Giese, R. H. *Icarus* **62**, 244-272 (1985).
8. Zook, H. A., Lange, G., Grün, E. & Fechtig, H. *Proc. lunar planet. Sci. Conf.* **15**, 965-966 (1984).

9. Clanton, U. S., Zook, H. A. & Schultz, R. A. *Proc. lunar planet. Sci. Conf.* **11**, 2261-2273 (1980).
10. McDonnell, J. A. M., Carey, W. C. & Dixon, D. G. *Nature* **309**, 237-240 (1984).
11. Naumann, R. J. NASA tech. Doc. No. 3717 (1966).
12. Brownlee, D. E., Tomandl, D. A., Hodge, P. W. & Horz, F. *Nature* **252**, 667-669 (1974).
13. McDonnell, J. A. M., Berg, O. E. & Richardson, F. F. *Planet. Space Sci.* **23**, 205-214 (1975).
14. Hoffman, H.-J., Fechtig, H., Grün, E. & Kissel, J. *Planet. Space Sci.* **23**, 215-224; 985-991 (1975).
15. McDonnell, J. A. M. in *Cosmic Dust* (ed. McDonnell, J. A. M.) 337-426 (Wiley, Chichester, 1978).

Heavy nitrogen in Bencubbin—a light-element isotopic anomaly in a stony-iron meteorite

I. A. Franchi, I. P. Wright & C. T. Pillinger

Planetary Sciences Unit, Department of Earth Sciences,
The Open University, Milton Keynes MK7 6AA, UK

Until now, light-element isotopic anomalies have been confined to primitive chondritic meteorites. Bencubbin, an unusual breccia found in Australia, comprising 60–75% metal with silicate and chondritic clasts¹⁻⁴ in a shock-welded matrix^{5,6}, has been shown⁷⁻⁹ to contain an unprecedented enrichment (by a factor of 2) of ^{15}N , presumably a nucleosynthetic product which escaped homogenization in the solar nebula. Here we show that virtually all the nitrogen in Bencubbin is enriched, with a maximum $\delta^{15}\text{N}$ of +1,033%. The nitrogen resides in two acid-resistant components, one more so than the other, which are present in different proportions in the metal and silicate. The more resistant component could be carbonaceous, with an unusually low C:N ratio and without an anomalous carbon isotopic composition ($\delta^{13}\text{C} \approx +25\%$). Another possible host is a chromium-rich sulphide, which would indicate processing in a supernova¹⁰.

Nitrogen isotopic measurements performed by Prombo and Clayton^{7,9} were restricted by sensitivity limitations to a comparison of the total nitrogen extracted from metal and silicate inclusions with that extracted from the parent sample. Interestingly, the bulk meteorite gave higher $\delta^{15}\text{N}$ values (+923 and +973%) than either of the major constituents (+829 and +489%, respectively), which has been interpreted⁷ as demonstrating the existence of a minor component with an even higher ^{15}N abundance. In an attempt to identify the heavy-nitrogen carrier, gases

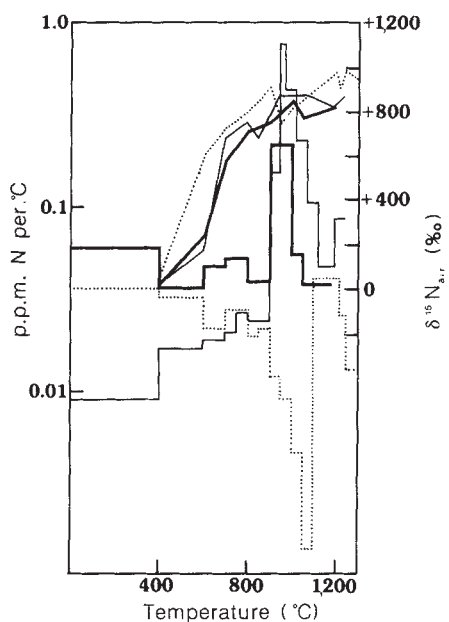


Fig. 1 Isotopic composition and release profile (histogram) of nitrogen liberated by stepped combustion from various components of Bencubbin: —, silicate clasts; ---, metal clasts; ····, glassy matrix. The nitrogen release pattern for Bencubbin is different from that of more common iron meteorites, which do not liberate any gas until 1,100 °C (I.A.F., unpublished data).

extracted from the sample in a stepped fashion have been analysed using a static mass spectrometer which is several orders of magnitude more sensitive than conventional dynamic machines and thus permits the isotopic analysis of minor components.

Figure 1 shows nitrogen concentration and isotopic data from stepped combustions of metal and silicate clasts. Both samples show evidence for atmospheric or biological contamination, characterized by $\delta^{15}\text{N}$ tending to 0‰ at temperatures <400 °C. If only gas extracted at temperatures >600 °C is considered, the silicate clast has a bulk $\delta^{15}\text{N}$ (+789‰) for 45 p.p.m. N, which approximates that measured for the iron-nickel metal (+868‰, 81 p.p.m. N). Neither of these isotopic values exceeds the whole rock $\delta^{15}\text{N}$ value of +973‰ measured by Prombo and Clayton^{7,9}, who also measured a much larger difference between the phases, possibly reflecting greater analytical uncertainty in their technique. More interesting is the similarity in the nitrogen release profiles and the $\delta^{15}\text{N}$ values for individual steps for the metal and silicate clasts. Both show two heavy-nitrogen components, the more abundant being released from 900 to 1,050 °C and characterized by the highest $\delta^{15}\text{N}$ value, +847‰ (silicate) or +888‰ (metal); for convenience, this nitrogen will be designated N_α . At lower temperatures (600–800 °C), there is a second release of nitrogen, with a slightly lower $\delta^{15}\text{N}$ value (+721‰, silicate and +745‰, metal); this nitrogen component will be referred to as N_β . The ratios of N_α to N_β are 3:1 for the silicate and 15:1 for the metal.

Figure 1 also shows data for a sample of matrix. Although the isotopic composition is similar to that of the metal and silicate clasts, the amount of nitrogen released is lower and peaks occur at distinctly different temperatures. Evidence for N_β exists in the temperature range 700–800 °C, but the yield of nitrogen declines to a distinct minimum in the expected temperature interval for N_α (900–1,050 °C), before a large release of nitrogen is observed at 1,100 °C. The N_α carrier phase could be 'protected' by glass from combustion until the onset of melting; this hypothesis *a priori* excludes matrix glass as the heavy-nitrogen host. If the nitrogen liberated from the matrix at >1,100 °C is N_α , this component cannot be more than a factor of ~2 in excess of N_β . It is interesting that the matrix has the

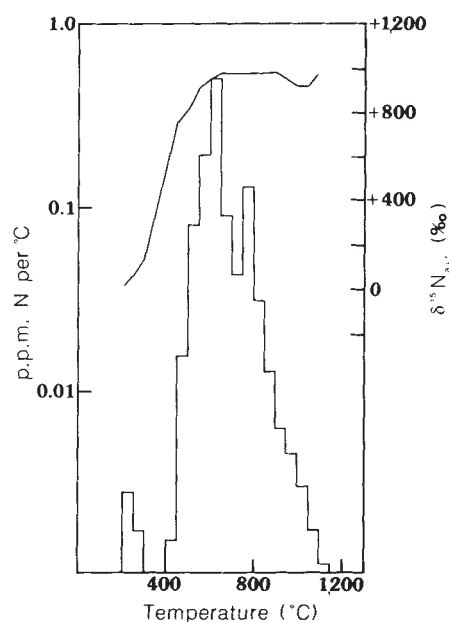


Fig. 2 Isotopic composition and release profile (histogram) of nitrogen liberated by stepped pyrolysis from a residue prepared from Bencubbin metal and matrix by treatment with 6 M HCl (see text for details).

lowest nitrogen content (14 p.p.m.), but the highest bulk and single-step $\delta^{15}\text{N}$ values (+879 and +996‰, 1,210–1,250 °C respectively) observed in our study. However, the $\delta^{15}\text{N}$ values are far too low for two-component mixing (small amounts of matrix adhering to the clasts) to account for the observed nitrogen isotope systematics of metal and silicate. Chondritic clasts may also be ruled out as primary carriers of isotopically heavy nitrogen. These were found to have only modest ^{15}N enrichments ($\delta^{15}\text{N} \leq +290$ ‰), possibly acquired during a secondary event.

To explore the nature of N_α and N_β , a 2-g sample of metal and matrix was treated with HCl and HF. The parent material, treated with 6 M HCl for 2 h at 70 °C, underwent a 92.1% weight loss. Pyrolysis of the residue in 50 °C steps (Fig. 2) produced 734 p.p.m. nitrogen, corresponding to 58 p.p.m. in the original sample, as might be expected from a mixture of metal (81 p.p.m. N) and matrix (14 p.p.m. N). Therefore, only a minor amount of the element was lost during dissolution. A small, isotopically light fraction must be missing, as the overall $\delta^{15}\text{N}$ of the residue is heavier than the starting material by >100‰. This light fraction could be ~10 p.p.m. nitrogen with an isotopic composition more typical¹¹ of meteoritic iron (~-80‰). Nitrogen release from the 6 M HCl residue is bimodal: the major constituent is liberated at 500–700 °C with a constant $\delta^{15}\text{N}$ value of +990 to +1,000‰, whereas the minor one, with a similar isotopic composition but seven times less abundant, appears suddenly at 800 °C. As the 2-g sample was mainly metal ($N_\alpha/N_\beta = 15$), with an indeterminate amount of matrix ($N_\alpha/N_\beta \approx 2$), the 7:1 relative abundance of the two nitrogen-containing species in the 6 M HCl residue is consistent with the lower-temperature component being N_α .

Stepped combustion of the 6 M HCl residue (Fig. 3) also gave a bimodal release ($N_\alpha/N_\beta = 5$), with $\delta^{15}\text{N}$ constant at +950 to +985‰, at lower temperatures than in the pyrolysis or combustion experiments on untreated metal, silicate or glass. Thus, the nitrogen carrier phases are actually degraded by hot oxygen (therefore they are not silicate), and are concealed in a substance which is soluble in 6 M HCl (metal or metal and glass).

Treatment of the 6 M HCl residue in harsher conditions (12 M HCl, 48 h at 25 °C) caused a weight loss (29.2%) and a reduction in the nitrogen content to 210 p.p.m. (equivalent to 12 p.p.m. in the parent). Figure 3 shows that N_α has decreased in the 12 M

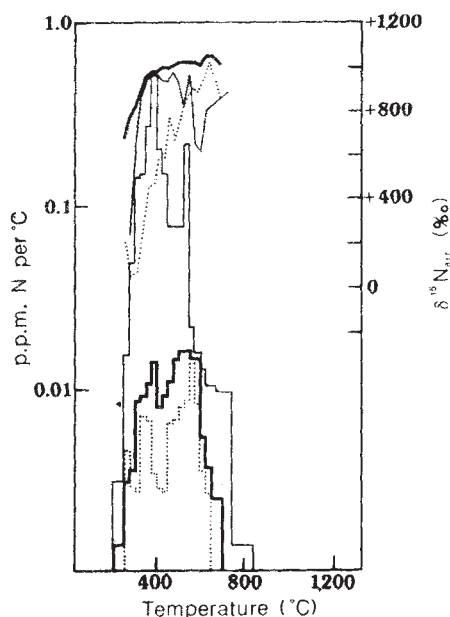


Fig. 3 Isotopic composition and release profile (histogram) of nitrogen liberated by stepped combustion from residues prepared from Bencubbin metal and matrix by treatment with: —, 6 M HCl; —, 12 M HCl; - - - - , 12 M HCl/48 M HF (see text for details).

HCl residue, so that N_β is now more abundant by a factor of 2.5. The observation that the N_α host phase is soluble in concentrated mineral acids is confirmed by treatment of the 6 M HCl residue with 1:1 12 M HCl/28 M HF for 72 h at 25 °C. The new residue represents only 0.03 wt% of the original specimen, but contains 8,047 p.p.m. nitrogen, which the combustion profile and isotopic values (Fig. 3) suggest is N_β (maximum $\delta^{15}\text{N} = +1,033\%$). The nitrogen observed in the HF/HCl residue is equivalent to ~3 p.p.m. of the original sample, which would be the expected bulk concentration of N_β for a mixture of metal and matrix. Thus, N_β is operationally defined to be almost completely stable in the most stringent mineral acid conditions.

Unfortunately, insufficient amounts of the acid residues were available for characterization by physical methods. The solubility of the N_α carrier in 12 M HCl and 28 M HF would appear to exclude the possibility that this material is carbonaceous. Carbon measurements show that there is a major release of carbon during stepped combustion of the 6 M HCl residue ($\delta^{13}\text{C} \approx 0\%$) at temperatures similar to the N_α release, but this carbon persists after treatment with stronger acids, whereas N_α is removed. Both the 6 M and 12 M HCl residues show a $\delta^{13}\text{C}$ maximum of +25% (only +6% for the HF/HCl residue, due to some contamination) in the temperature range 500–600 °C, which corresponds to the release of N_β . The amount of carbon in the HF/HCl residue burning between 450 and 700 °C is <5,000 p.p.m.; therefore, if the +25% carbon phase is associated with N_β , the C:N ratio is <1. Allowing all the carbon in the HF/HCl residue to be the host for N_β would not give a C:N ratio >4. A $\delta^{13}\text{C}$ value of +25% is hardly anomalous, compared with the $\delta^{15}\text{N}$ of ~+1,033%, but it is unusual for carbon burning at <600 °C in meteorite acid residues.

The 6 M HCl residue yielded 0.84 wt% S ($\delta^{34}\text{S} = -1.0\%$) over the temperature range 300–500 °C, suggesting that sulphur is present as a sulphide¹². This release temperature coincides with that of the N_α and N_β components. Although most sulphides are soluble in HCl, chromium-rich sulphides, up to 30% Cr, have been observed in Bencubbin^{5,6}. Based on the stability of daubreelite (FeCr_2S_4), also reported in Bencubbin¹³, these could have properties similar to the N_α or N_β host.

The $\delta^{15}\text{N}$ values >1,000‰ for N_α and N_β cannot be explained by normal, mass-dependent fractionation effects such as degassing. Photochemical, non-mass-dependent processes, such as can be invoked for oxygen¹⁴, may be possible, but it will be difficult to demonstrate their involvement. On the other hand, nucleosynthesis of ^{15}N -enriched nitrogen in a particular stellar environment such as a nova⁹ or supernova can provide an explanation,

especially if isotopic effects for the other elements can be recognized. Unfortunately, no unusual noble gas isotopic compositions (except a possible sub-solar component) have been recognized in Bencubbin^{15,16}, which could shed light on the origin of the nitrogen anomaly.

The possibility that sulphide minerals are the carriers is fascinating, as Cr- and Ti-sulphides might be produced in supernovae¹⁰, and the hot CNO cycle operating in such a star, evolved to the silicon-burning stage, could produce nitrogen with extreme $^{15}\text{N}/^{14}\text{N}$ ratios¹⁷. Indeed, if the cycle occurs in a helium-rich zone, near-solar $^{13}\text{C}/^{12}\text{C}$ ratios can be produced in association with high ^{15}N abundances. It is not clear, however, why the chromium-rich sulphides should be located at the margins of metal grains⁵.

From our experiments it seems likely that almost all the nitrogen in Bencubbin is enriched in ^{15}N , rather than a small proportion having an extreme $\delta^{15}\text{N}$ value being diluted by isotopically normal gas⁹. The similarity in the $\delta^{15}\text{N}$ values measured for N_α and N_β argue in favour of nitrogen from a single source being trapped in coexisting, possibly related, carrier phases. The existence of both carriers, in different proportions and absolute abundance, in metal and silicate clasts is enigmatic. Their existence in the matrix can be explained if the matrix consists of remobilized metal and silicate clasts which have been partially degassed during shock-melting. The release of the N_α host from its location within the meteorite by 6 M HCl suggests that it is accommodated in metal, and supports the argument that Bencubbin metal could be a nebular condensate⁵ which was seeded by pre-existing grains.

We thank R. Burgess for sulphur measurements and Dr Robert Hutchison (British Museum) for samples and continued interest in this project, which was supported by SERC.

Received 10 March; accepted 26 June 1986.

1. Simpson, E. S. & Murray, D. G. *Mineralog. Mag.* **23**, 33–37 (1932).
2. Lovering, J. S. *Researches on Meteorites* (ed. Moore, C. B.) 179–197 (Wiley, New York, 1962).
3. McCall, G. J. H. *Mineralog. Mag.* **36**, 726–739 (1968).
4. Mason, B. & Nelen, J. *Geochim. cosmochim. Acta* **32**, 661–664 (1968).
5. Newsom, H. E. & Drake, M. J. *Geochim. cosmochim. Acta* **43**, 689–707 (1979).
6. Hutchison, R. *Lunar planet. Sci.* **XVII**, 374–375 (1986).
7. Prombo, C. A. & Clayton, R. N. *Lunar planet. Sci.* **16**, 679–680 (1985).
8. Franchi, I. A. et al. *Meteoritics* **20**, 645–646 (1985).
9. Prombo, C. A. & Clayton, R. N. *Science* **230**, 935–937 (1985).
10. Clayton, D. D. & Ramadurai, S. *Nature* **265**, 427–428 (1977).
11. Prombo, C. A. & Clayton, R. N. *Meteoritics* **18**, 377–379 (1983).
12. Burgess, R. et al. *Lunar planet. Sci.* **16**, 99–100 (1985).
13. Ramdohr, P. *The Opaque Minerals in Stony Meteorites* (Elsevier, Berlin, 1973).
14. Thiemens, M. H. & Heidenreich, J. E. *Science* **219**, 1073–1075 (1983).
15. Begemann, F. et al. *Geochim. cosmochim. Acta* **40**, 353–368 (1976).
16. Lewis, R. S. *Meteoritics* **20**, 698–699 (1985).
17. Audouze, J. & Vauclair, S. *Introduction to Nuclear Astrophysics* (Reidel, Dordrecht, 1980).

Actinide enrichment in marine aerosols

M. I. Walker*, W. A. McKay†, N. J. Pattenden
& P. S. Liss*

* School of Environmental Sciences, University of East Anglia,
Norwich NR4 7TJ, UK

† Environmental and Medical Sciences Division,
UKAEA Harwell Laboratory, Oxfordshire OX11 0RA, UK

Previous work¹⁻⁴ has shown that a small fraction of the plutonium and americium discharged into the Irish Sea from the nuclear fuels reprocessing plant at Sellafield, Cumbria (United Kingdom) has been transferred into the atmosphere and returned to land. Bubble scavenging in the water column, coupled with droplet ejection from bubbles bursting at the surface, may be involved in the transfer mechanism; other studies⁵⁻⁷ have shown this to be true for heavy metals. Laboratory measurements^{8,9} have demonstrated aerosol enrichments for radionuclides. Here we describe investigations of the sea-to-air transfer of Pu and Am. Initial results indicate that, relative to their concentrations in bulk sea water, these actinides are enriched in artificially generated marine aerosols by factors of 30–600. We believe this to be the first field study which demonstrates that bursting bubbles can contribute to enhanced actinide levels in marine air.

Figure 1 shows schematically the equipment which we used to investigate the sea-to-air transfer of radionuclides by a bubble-bursting mechanism. Bubbles (200–1,000 μm diameter) are generated by passing air through a glass frit below the sea surface; when they burst, the droplets formed are drawn onto a Whatman 541 filter by a commercial high-volume pump. The collection apparatus is carried between the hulls of a catamaran. With the bubble source operating, the equipment can collect up to 15 ml h^{-1} of droplets as marine aerosol. Blanks, that is, samples collected without bubble generation, are equivalent to $0.05\text{--}0.1 \text{ ml h}^{-1}$ (these volumes are expressed as equivalent volumes of sea water collected on the filter, based on the sodium content of filters and seawater samples). All blank actinide levels were below the detection limit, implying that the atmospheric contribution to the lowest-activity samples was $<20\%$.

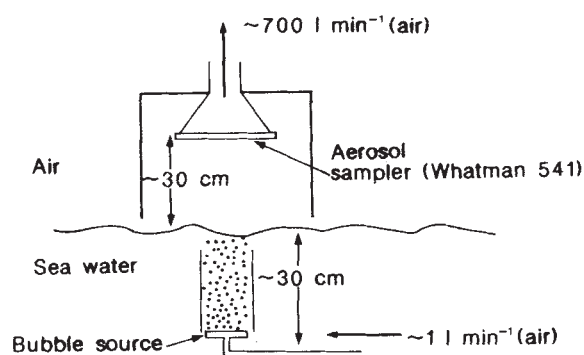


Fig. 1 The Harwell bubble-burst aerosol sampler.

The equipment has been used at sea on two occasions (April 1984) within 2 km of the Sellafield marine discharge pipeline and on three occasions $\sim 10 \text{ km}$ to the north-west. Aerosol samples with and without the bubble source operating were taken, together with bulk seawater samples. Seawater samples were filtered through membrane filters with $0.45 \mu\text{m}$ pore size. On 26 April 1984 a sea-surface microlayer sample $\sim 150 \mu\text{m}$ thick was also collected, using a Garrett screen¹⁰, and filtered.

Samples were analysed for actinides (by α -spectrometry after radiochemical separation¹¹), for fission products (by γ -spectrometry) and for stable sodium (by inductively coupled plasma optical emission spectroscopy (ICP)).

Table 1 lists ^{238}Pu , $^{239+240}\text{Pu}$, ^{241}Am and ^{137}Cs activities in the samples of aerosol (normalized to a volume of 1 litre by relating aerosol to bulk-water sodium content), filtered sea water and suspended particulate. Errors quoted are 2 standard deviations, based on counting statistics.

Enrichment of a material X in marine aerosols and microlayers is best expressed in terms of an enrichment factor, $\text{EF}(\text{X}) = [\text{X}]_{\text{aerosol or microlayer}} / [\text{X}]_{\text{bulk sea water}}$ where [X] is the normalized concentration of X. These have been calculated in Table 2 for aerosol and microlayer samples, for total, particulate and filtrate fractions, as appropriate.

From the enrichment factors for actinides in the microlayer, it appears that there is no enrichment in the filtrate, any

Table 1 Seawater, aerosol and microlayer radionuclide activities

Date	Sample	^{238}Pu (mBq l ⁻¹)	$^{239+240}\text{Pu}$ (mBq l ⁻¹)	^{241}Am (mBq l ⁻¹)	^{137}Cs (Bq l ⁻¹)
26.04.84	Seawater filtrate	9.2 ± 0.7	28.1 ± 1.5	6.3 ± 0.7	11.5 ± 2.2
	Seawater particulate	1.7 ± 0.4	5.3 ± 0.8	5.0 ± 0.6	0.03 ± 0.02
	Seawater total	10.9 ± 0.8	33.4 ± 1.7	11.3 ± 1.0	11.5 ± 2.2
	Aerosol	436 ± 92	$1,512 \pm 218$	$1,045 \pm 247$	15.9 ± 8.6
	Microlayer filtrate	10.0 ± 2.2	30.3 ± 4.1	0.7 ± 0.7	13.6 ± 2.5
	Microlayer particulate	33.7 ± 5.5	104 ± 9.2	42.9 ± 6.7	0.2 ± 0.1
27.04.84	Microlayer total	43.7 ± 5.9	134 ± 10.1	43.6 ± 6.7	13.8 ± 2.5
	Seawater filtrate	4.4 ± 0.7	16.3 ± 1.5	3.0 ± 0.9	7.0 ± 2.5
	Seawater particulate	2.1 ± 0.4	6.5 ± 0.9	4.0 ± 0.5	0.01 ± 0.01
	Seawater total	6.6 ± 0.8	22.8 ± 1.7	7.0 ± 0.9	7.0 ± 2.5
	Aerosol	488 ± 95	$1,540 \pm 230$	780 ± 195	25.0 ± 12.4
	Seawater filtrate	4.8 ± 0.7	18.9 ± 1.5	2.6 ± 0.7	16.8 ± 1.8
23.08.84	Seawater particulate	2.8 ± 0.4	9.7 ± 0.9	7.8 ± 0.8	0.02 ± 0.003
	Seawater total	7.6 ± 0.8	28.6 ± 1.7	10.4 ± 1.1	16.9 ± 1.8
	Aerosol	213 ± 98	804 ± 188	$1,040 \pm 222$	25.7 ± 9.7
	Seawater filtrate	1.0 ± 0.1	4.7 ± 0.2	0.4 ± 0.1	2.8 ± 0.9
	Seawater particulate	0.9 ± 0.1	3.8 ± 0.3	3.9 ± 0.4	0.01 ± 0.01
	Seawater total	1.9 ± 0.1	8.5 ± 0.4	4.3 ± 0.4	2.8 ± 0.9
11.09.85	Aerosol	572 ± 352	$2,970 \pm 807$	$2,516 \pm 580$	<31
	Seawater filtrate	1.4 ± 0.1	5.4 ± 0.3	1.2 ± 0.3	0.9 ± 0.4
	Seawater particulate	1.9 ± 0.3	7.8 ± 0.9	8.9 ± 1.1	0.01 ± 0.01
	Seawater total	3.3 ± 0.3	13.2 ± 1.0	10.1 ± 1.1	0.9 ± 0.4
	Aerosol	522 ± 270	955 ± 163	$1,040 \pm 397$	<13

Particulate activities represent the activity associated with suspended particulate material per litre of sea water or microlayer filtered; aerosol activities are normalized to a volume of 1 litre of sea water.

Table 2 Enrichment factors

Date	Sample		^{238}Pu	$^{239+240}\text{Pu}$	^{241}Am	^{137}Cs	Al
26.4.84	Aerosol	Total	40 ± 9	45 ± 7	92 ± 23	1.4 ± 0.8	220 ± 50
		Particulate	260 ± 86	283 ± 61	207 ± 56		
	Microlayer	Total	4.1 ± 0.6	4.0 ± 0.4	3.9 ± 0.7	1.2 ± 0.3	
		Filtrate	1.1 ± 0.2	1.1 ± 0.2	0.1 ± 0.1*	1.2 ± 0.3	
27.4.84	Aerosol	Total	20 ± 6.8	20 ± 3.4	8.3 ± 1.4	7 ± 5	3.6 ± 2.2
		Particulate	74 ± 17	68 ± 11	112 ± 31		
	Aerosol	Total	228 ± 62	235 ± 48	194 ± 54	1.5 ± 0.6	
		Particulate	28 ± 13	28 ± 7	100 ± 24		
23.8.84	Aerosol	Total	77 ± 37	82 ± 21	132 ± 31		660 ± 150
		Particulate	291 ± 180	347 ± 96	583 ± 145	<11	
	Aerosol	Total	616 ± 388	776 ± 222	638 ± 160		
		Particulate	158 ± 83	73 ± 14	102 ± 39	<14	
16.10.85	Aerosol	Total	269 ± 146	124 ± 26	116 ± 45		230 ± 60
		Particulate					

Aerosol activities are divided by total (solution plus particulate) seawater activities to give total enrichment factors (EF). For explanation of particulate EFs see text. Microlayer EFs are calculated by dividing the microlayer activities by the corresponding seawater activity.

* This result is anomalous because the ^{241}Am activity in the microlayer filtrate (Table 1) is very low. As the $^{241}\text{Am}/^{239+240}\text{Pu}$ ratio in this sample appears to be much smaller than all other measurements of this ratio by ourselves and others, we conclude that the ^{241}Am value is probably wrong for unknown reasons, and that the apparent depletion of ^{241}Am is not real.

enrichment being a result of some particulate process. By comparison of suspended particulate loads, the surface microlayer sample appears to be enriched in particulate material relative to bulk sea water.

Assuming that, as in the microlayer, the solution component of the aerosol is not enriched, then from the equivalent volumes of aerosol collected it can be calculated that <3% of the actinides in the aerosol are derived from the solution phase. Some aerosol and suspended particulate samples were analysed for aluminium, by ICP, as an indicator of clay-mineral particulate material; enrichment factors for aerosol Al relative to suspended particulate Al are given in Table 2. If it is assumed that all the actinides in the aerosol are transferred from sea to air in particulate form, then EF values can be calculated relative to the seawater particulate fraction, to give the particulate EFs in Table 2. The actinide particulate EFs are comparable to the aluminium EFs.

The previous measurements¹⁻⁴ have shown that although the fission product ^{137}Cs can be detected in atmospheric and terrestrial studies, there is no evidence that it is enriched in marine aerosol. This is confirmed in the present work. Unlike the actinides, it behaves conservatively in sea water¹², and hence is not strongly associated with marine particulates.

Sea-surface microlayers have been found to be enriched by 1-2 orders of magnitude in heavy metals¹³ such as Fe, Cu and Zn, and by factors of 1-8 in radionuclides¹⁴ such as ^{210}Po and ^{210}Pb . Piotrowicz *et al.*⁵ and Weisel *et al.*⁶, using the Bubble Interfacial Microlayer Sampler¹⁵ (BIMS), found enrichment factors of 200-800 for Cu, 5,000 for Pb, <100-10,000 for Fe and 200-20,000 for Zn in the North Atlantic. The EF values for Zn obtained with the BIMS increased with increasing depth of bubbling, suggesting that scavenging is important. Pattenden *et al.*⁷, working in the North Sea, found EFs of up to 500 for Zn and 2,400 for Pb.

In the Irish Sea the sea-surface microlayer and artificially generated marine aerosols are enriched in actinides. The particulate phase is thought to be largely responsible. It seems likely that the actinides in the aerosol, observed to be enriched, are tracers of a natural effect which can produce large enrichments of clay-mineral particles from the water column. High particulate loads and bubble fluxes in surface waters may increase the size of the effect. As Blanchard¹⁶ has reported a sea-to-land effect for bacteria, it seems that this phenomenon may well be of wider significance. Trace metals and organic materials could also be transferred from sea to land by this mechanism, and act as harmful pollutants or nutrients.

The transport to land by onshore winds of actinides in marine aerosols is dependent on the residence time, and hence the size,

of the aerosol droplets in near-ground air, as is the case for sea salt. Thus, the atmospheric dispersion and deposition of actinides is likely to be similar to that of sea salt¹⁷. While the bubble-bursting mechanism can generate droplets of 0.5 to >400 μm diameter^{18,19}, only those smaller than 100 μm are likely to travel more than a few hundred metres.

The radiation dose from sea-to-land transfer, arising predominantly from large droplets produced in the surf zone²⁰, has been shown to be negligible (<1%) in comparison to that received by the critical group from seafood consumption^{21,22}. However, the small droplets generated by the bubble-bursting mechanism may make a measurable contribution to levels of Pu in air (airborne concentrations <1% of ICRP (International Commission on Radiation Protection) limits for members of the public) and deposition¹ near the Cumbrian coast, and may be responsible for the trace quantities of Sellafield-derived Pu which have been reported²³ at distances of up to 60 km from the plant. This work is being continued to investigate mechanisms in more detail, to examine size distributions of actinides in the aerosol and to calculate sea-to-land fluxes.

This work has been supported by the UK Atomic Energy Authority, the UK Department of the Environment, the Commission of the European Communities and the US NSF (as part of the SEAREX program). We thank our colleagues J. D. Eakins, L. Salmon, G. N. J. Lewis and A. G. Morton for analytical work, and R. S. Cambray, J. R. Branson and K. Playford for general assistance.

Received 10 February; accepted 20 June 1986.

- Pattenden, N. J., Cambray, R. S., Playford, K., Eakins, J. D. & Fisher, E. M. R. in *Impacts of Radionuclide Release into the Marine Environment* (IAEA SM-248/138), 201-221 (International Atomic Energy Agency, Vienna, 1981).
- Cambray, R. S. & Eakins, J. D. *Nature* **300**, 46-48 (1982).
- Peirson, D. H., Cambray, R. S., Cawse, P. A., Eakins, J. D. & Pattenden, N. J. *Nature* **300**, 27-31 (1982).
- Pattenden, N. J., Cambray, R. S. & Eakins, J. D. in *Ecological Aspects of Radionuclide Release* (eds Coughtrey, P. J., Bell, J. N. B. & Roberts, T. M.) 259-271 (Blackwell, Oxford, 1983).
- Piotrowicz, S. R., Duce, R. A., Fasching, J. L. & Weisel, C. P. *Mar. Chem.* **7**, 307-324 (1979).
- Weisel, C. P., Duce, R. A., Fasching, J. L. & Heaton, R. W. *J. geophys. Res.* **89**, 11607-11618 (1984).
- Pattenden, N. J., Cambray, R. S. & Playford, K. *Geochim. cosmochim. Acta* **45**, 93-100 (1981).
- Belot, Y., Caput, C. & Gauthier, D. *Atmos. Envir.* **16**, 1463-1466 (1982).
- Garland, J. A. & Chadwick, R. C. *Health Phys.* **41**, 279-283 (1981).
- Garrett, W. D. *Limnol. Oceanogr.* **10**, 602-605 (1965).
- Eakins, J. D. *Nucl. Instrum. Meth.* **223**, 194-199 (1984).
- Hetherington, J. A., Jefferies, D. F. & Lovett, M. B. in *Impacts of Nuclear Release into the Aquatic Environment* (IAEA SM-198/29) 193-210 (International Atomic Energy Agency, Vienna, 1975).
- Hardy, J. T., Apts, C. W., Crecelius, E. A. & Bloom, N. S. *Estuar. Coast. Shelf Sci.* **20**, 299-312 (1985).
- Bacon, M. P. & Elzerman, A. W. *Nature* **284**, 332-334 (1980).
- Fasching, J. L., Courant, R. A., Duce, R. A. & Piotrowicz, S. R. *J. Rech. atmos.* **8**, 649-652 (1974).

16. Blanchard, D. C. in *Air-Sea Exchange of Gases and Particles* (eds Liss, P. S. & Stinn, W. G. N.) 407-454 (Reidel, Dordrecht, 1983).
17. Rossnecht, G. F., Elliott, W. P. & Ramsey, F. L. *J. appl. Met.* **12**, 825-830 (1973).
18. Exton, H. J., Latham, J., Park, P. M., Perry, S. J. & Smith, M. H. *Q. Jl R. met. Soc.* **111**, 817-837 (1985).
19. Wu, J., Murray, J. J. & Lai, R. J. *J. geophys. Res.* **89**, 8163-8169 (1984).
20. Eakins, J. D. & Lally, A. E. *Sci. tot. Envir.* **35**, 23-32 (1984).
21. Wade, B. O., Eggleton, A. E. J. & Peirson, D. H. *Studies of Environmental Radioactivity in Cumbria*, Pt 7 (AERE Rep. No. R11743) (HMSO, London, 1985).
22. Linsley, G. S., Dionion, J., Simmonds, J. R. & Burgess, J. *National Radiological Protection Board Rep. No. R-170* (HMSO, London, 1984).
23. Eakins, J. D. *et al. J. enviro. Radioact.* **1**, 87-105 (1984).

Seismic reflectors, conductivity, water and stress in the continental crust

D. Ian Gough

Department of Physics, University of Alberta, Edmonton, Canada T6G 2J1

The uppermost 10-15 km of the Earth's continental crust differs in several geophysical properties from the lower crust. The upper crust is electrically resistive, seismically transparent, contains nearly all intracontinental earthquake hypocentres and responds to stress elastically, with brittle fracture. The lower crust is electrically conductive, contains many seismic reflectors, is aseismic and shows ductile response to stress. I show here that the characteristics of both regions can be explained if the entire crust contains saline water, in separated cavities in the compressively stressed rocks of the upper crust, but in the lower crust forming an interconnected film on the crystal surfaces. The model may apply principally to tectonically active parts of the continental crust; beneath shields, the lower crust may be dry.

Two remarkable features of the continental crust, illustrated in Fig. 1, have been recognized in recent geophysical results from western Europe and North America. Deep seismic profiles¹⁻⁴ show many short, flat or slightly arcuate reflectors below ~10-15 km depth, in contrast to an upper crust which is much more transparent, below any superficial sediments. Although there are reflectors in many parts of the upper crust, the many profiles now available demonstrate its relative transparency at frequencies of a few hertz. The nature of the lower-crustal reflectors is speculative, but the seismic amplitudes favour multiple boundaries giving constructive interference rather than single interfaces⁵⁻⁷. Layered rocks such as mylonites, produced by prolonged shearing, or layered intrusions could provide such multiple interfaces.

The second feature is an electrically conductive layer in the same depth range as the top of the lower-crustal seismic reflectors⁸⁻¹⁰. Resistivities typically in the range 1-50 Ω m characterize this layer, in contrast to values of 10^3 - 10^5 Ω m in the upper crust. The low resistivities at mid-crustal depths could be associated with carbon or with metal sulphides, but there is good reason to expect water in fractures and pores throughout the upper crust¹¹, so that electrolytic aqueous solutions in interconnected cavities are the most likely widespread good conductors in the continental crust. The fluid will be called brine, for brevity.

If the upper crust contains brine, and wherever the basement is seen it is densely fractured, it is reasonable to ask why the upper crust is resistive before going on to ask why there is a conductive layer at mid-crustal depths. *A priori*, one might expect a conductive upper crust. Similarly one might ask why the upper crust is transparent, before asking why the lower crust is reflective.

A possible explanation of the resistive upper crust lies in its stress field. In the upper crust, in general, one principal stress is sub-vertical, the other two horizontal. The vertical principal stress is within ~20% of the lithostatic pressure¹² $\rho_r gz$ (where

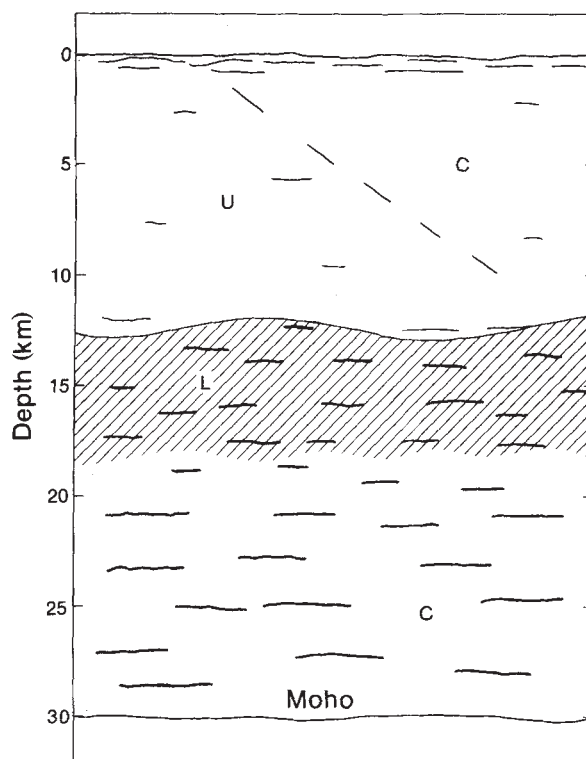


Fig. 1 The continental crust; horizontal lines represent seismic reflectors. UC, Upper crust with near-surface sediments and one through-going fault, few seismic reflectors and high resistivity. LC, Lower crust with many reflectors, low resistivity shown by shading. The top of the conductive zone is better defined than its thickness.

ρ_r is the density of the overlying rock, z is the depth and g is the acceleration due to gravity), so water at hydrostatic pressure $\rho_w gz$ cannot hold horizontal fractures open. If the lesser horizontal principal stress S_h is large enough to close fractures and isolate the brine in closed cavities, the bulk resistivity will be only slightly below that of the dry rock, and thus relatively high. Both North America east of the Rocky Mountains, and western Europe, have upper crusts under large compressive horizontal stresses of tectonic origin¹³⁻¹⁶ and the geophysical data discussed here refer mainly to these regions. In both, the greater horizontal principal stress, S_H , exceeds the vertical lithostatic stress, and the lesser horizontal compression, S_h , is comparable to the vertical stress. Most fractures will be held closed by the horizontal compression, and the brine will be separated in unconnected cavities. If the horizontal stress remains compressive over geologically short times (thousands of years), many fractures will fill with minerals such as quartz or epidote. Only large, frequently active faults, such as through-going thrusts, might be expected to remain open. An upper crust under horizontal compression should thus be electrically resistive and also seismically transparent, except at major through-going faults. This is the situation observed².

In regions of strike-slip faulting, such as the San Andreas fault system of California, where the relative motion involves no underthrust cold lithosphere, earthquake foci are essentially confined to the upper crust¹⁷. This is true for most earthquakes in continental interiors, so that failure is clearly brittle in the top 15 km and ductile below¹⁷. Nearly elastic response to stress is to be expected in the upper crust until brittle fracture occurs. The high electrical resistivity implies that the fractures are closed where S_h is large enough. This will be the case except in regions of extensional stress. The least horizontal stress possible in rock is given by the elastic response to the vertical load (the lithostatic pressure), $\rho_r gz$ at depth z . This least horizontal pressure is

$\rho_r g z / (1 - n)$ in rock of Poisson's ratio n , or $\rho_r g z / 3$ for $n = 0.25$. As water has density about one third of that of crustal rock, water-filled vertical cracks can remain open at all depths at which the response is elastic, in the absence of tectonic horizontal compression—that is, in a rifted region. Low and strongly anisotropic resistivities are to be expected in such regions, as is observed¹⁸. If the lesser horizontal stress, S_h , exceeds half the vertical stress, the upper crustal resistivity should be high and approximately isotropic. The limited available data indicate that the latter situation is the norm: where the stress field is known in continental rocks, all three principal stresses exceed the hydrostatic pressure.

A rather relaxed stress regime exists in the Kaapvaal craton of South Africa, where the upper-crustal stress field is well known from measurements in deep mines¹². At depths of <500 m the horizontal pressures exceed the vertical. At greater depths the vertical principal stress is greatest, but the horizontal pressures are not much smaller and certainly exceed the hydrostatic pressure at all depths. As expected, the upper crust is highly resistive¹⁹.

An orthodox view of the lower crust is that it is a very dry region, of thoroughly dehydrated granulitic composition. This view is based largely on studies of the chemistry of rocks which have passed through dehydration and rehydration reactions during burial to mid-crustal depths, followed by uplift^{20,21}. Such rocks commonly show a very dry mineralogy with limited and uneven rehydration. These facts are not in conflict with an excess of a few per cent of free water in the lower crust, if water is lost during uplift and before rehydration. The extraction of metamorphic fluid is poorly understood, but probably occurs during uplift by the development of microcracks as the fluid expands more than the rock (W. S. Fyfe, personal communication). Rock cores must be compressed for reliable laboratory measurements of seismic velocities²², because they contain many microscopic cracks. These can be explained by expansion and release of fluid during uplift.

With a dry lower crust it would be difficult to account for the ductile rheology there: one would expect earthquake hypocentres in the lower crust, not the aseismic sliding which is certainly proceeding on transform faults like the San Andreas. As intracontinental earthquakes occur mainly in tectonically active regions, the evidence for a ductile lower crust applies mainly there. The lower crust beneath shields may be of the dry granulite type. Elsewhere, the geophysical evidence invites examination of the geochemistry and mineralogy of a lower crust with a small amount of excess free water.

Fyfe²³ has calculated that a water mass equal to that of the oceans is recycled by subduction into the mantle in $\leq 10^9$ yr, and has discussed the release of large quantities of water into the basal crust in subduction zones and continental collisions²⁴. Geophysical evidence for the subduction of wet sediments has recently been published²⁵. It is thus reasonable to suppose that much of the lower continental crust may be a saturated environment, with excess water in equilibrium with a hydrated mineralogy unfamiliar at the surface. The suggestion that wet granitic rocks in the lower crust could account for high electrical conductivity there is not new²⁶. A wet lower continental crust may prove to be a feature of plate tectonics, at least in active regions.

Even pure water has an ion density in the middle crust which is ≥ 5 orders of magnitude greater than at the surface (ref. 11, p. 26), and lower-crustal water may be a strong brine of halides as well as HCl (refs 11, 27, 28). The low bulk resistivity can be explained if the water is not contained in pockets, but coats mineral grains to form a conductive mesh. Such material might well deform in a ductile manner, with flow in response to small stress differences. In a general way, this seems promising for shear deformation producing mylonitic seismic reflectors. In terms of crustal water, the upper crust may be characterized by containment of the brine in closed cavities. In a transitional

depth range ($P \approx 3\text{--}4$ kbar, $T \approx 200\text{--}300^\circ\text{C}$) the cavities may break down, so that the brine forms an intergranular film in the lower crust. The breakdown of cavities filled with brine at lithostatic pressure is unlikely to occur by simple mechanical failure in response to a stress difference. More probably, chemical reactions are involved, with accompanying mechanical failure. Hydrated minerals may contribute to the electrical conductivity of the lower crust^{29,30}, in addition to the intergranular film of residual brine.

I thank W. S. Fyfe for stimulating discussions and B. W. D. Yardley for a useful review. My research is supported by the Natural Sciences and Engineering Research Council of Canada.

Received 11 March; accepted 6 June 1986.

1. Cook, F. A. & Oliver, J. E. *Am. J. Sci.* **281**, 993–1008, 1981.
2. Matthews, D. H. & Cheadle, M. J. *Am. geophys. Un. Geodyn. Ser.* **13**, 5–19 (1986).
3. Meissner, R. & Wever, T. *Am. geophys. Un. Geodyn. Ser.* **13**, 31–42 (1986).
4. Blundell, D. J. & Raynaud, B. *Am. geophys. Un. Geodyn. Ser.* **13**, 287–295 (1986).
5. Fuchs, K. Z. *Geophys.* **35**, 133–149 (1969).
6. O'Doherty, R. F. & Anstey, N. A. *Geophys. Prospect.* **19**, 430–458 (1971).
7. Fountain, D. M., Hurich, C. A. & Smithson, S. B. *Geology* **12**, 195–198 (1984).
8. Jödicke, H., Untiedt, J., Olgemann, W., Schulte, L. & Wagenitz, V. in *Plateau Uplift* (eds Fuchs, K. et al.) (Springer, Berlin, 1983).
9. Thompson, B. G., Nekut, A. & Kuckes, A. F. *J. geophys. Res.* **88**, 9461–9473 (1981).
10. Haak, V. & Hutton, V. R. S. *Pap. presented at Meet. on Lower Continental Crust* (Royal Astronomical Society, London, 1984).
11. Fyfe, W. S., Price, N. J. & Thompson, A. B. *Fluids in the Earth's Crust* (Elsevier, Amsterdam, 1978).
12. McGarr, A. & Gay, N. C. A. *Rev. Earth planet. Sci.* **6**, 405–436 (1978).
13. Zoback, M. L. & Zoback, M. D. *J. geophys. Res.* **85**, 6113–6156 (1980).
14. Gough, D. I., Fordjor, C. K. & Bell, J. S. *Nature* **305**, 619–621 (1983).
15. Ahorner, L. *Tectonophysics* **29**, 233–249 (1975).
16. Greiner, G. & Illies, J. H. *Pure Appl. Geophys.* **115**, 11–26 (1977).
17. Chen, W.-P. & Molnar, P. *J. geophys. Res.* **88**, 4183–4214 (1983).
18. Rooney, D. & Hutton, V. R. S. *Geophys. J. R. astr. Soc.* **51**, 91–119 (1977).
19. van Zijl, J. S. V. *Trans. geol. Soc. S. Afr.* **81**, 129–142 (1978).
20. Fyfe, W. S. *Phil. Trans. R. Soc. A* **283**, 221–228 (1976).
21. Yardley, B. W. D. *Geology* **9**, 405–408 (1981).
22. Birch, F. *J. geophys. Res.* **65**, 1083–1102 (1960); **66**, 2199–2224 (1961).
23. Fyfe, W. S. *Am. geophys. Un. Geodyn. Ser.* **5**, 82–86 (1981).
24. Fyfe, W. S. *Am. geophys. Un. Geodyn. Ser.* **14**, 33–39 (1986).
25. Kurtz, R. D., DeLaurier, J. M. & Gupta, J. C. *Nature* **321**, 596–599 (1986).
26. Hyndman, R. D. & Hyndman, D. W. *Earth planet. Sci. Lett.* **4**, 427–432 (1968).
27. Fyfe, W. S. *Tectonophysics* **119**, 29–36 (1985).
28. Tromsdorff, V., Skippen, G. & Ulmer, P. *Contr. Miner. Petrol.* **89**, 24–29 (1985).
29. Stesky, R. M. & Brace, W. F. *J. geophys. Res.* **78**, 7614–7621 (1973).
30. de Beer, J. H., van Zijl, J. S. V. & Gough, D. I. *Tectonophysics* **83**, 205–225 (1982).

Large Variscan overthrusts beneath the Paris Basin

M. Cazes*, A. Masclet†, X. Torrelles*, Ch. Boist†, X. Damotte†, Ph. Matte‡, X. Raoults§, V. N. Pham||, A. Hirn¶ & X. Galdeano¶

* Société Nationale Elf-Aquitaine, Tour Elf, Cedex 45, 92078 Paris La Défense, France

† Institut Français du Pétrole, BP 311, 92506 Reuil Malmaison Cedex, France

‡ Université des Sciences et Technologie Languedoc, Place E. Bataillon, 34060 Montpellier Cedex, France

§ Universités des Sciences et Technologie Lille, Laboratoire de Géologie, BP 35, 59650 Villeneuve d'Ascq, France

|| Laboratoire de Géophysique Appliquée et Structurale, Rue Doyen Roubaud, BP 40, 54501 Vandœuvre les Nancy, France

¶ Institut de Physique du Globe Paris, Université Pierre et Marie Curie, 4, place Jussieu, 75230 Paris Cedex 05, France

The ECORS project is devoted to the deep geophysical investigation of the continental and oceanic crust in France through the study of fundamental problems such as the formation of basins, mountain belts and continental margins. The first of the profiles completed is the 'Nord de la France', which is mainly devoted to the survey of the major units of the Northern Hercynian Belt and of the deep crust below the Mesozoic Paris basin. The deep geophysical investigations undertaken on this profile have led to

the acquisition of diverse information, such as vibroseismic, wide-angle reflection, refraction and magnetotelluric data. The main results presented here show evidence of large overthrusts in the Variscan chain which is currently buried beneath the thin Mesozoic cover of the Paris basin. They show a differentiation of structural units in the upper crust, their link with the lower crust, and the depth and geometry of the Moho. They also give rise to some speculations regarding the nature of various gravimetric and magnetic anomalies such as the large 'anomalie magnétique du Bassin de Paris' (AMBP).

The Nord de la France profile extends for 230 km between the towns of Cambrai and Dreux. Figure 1 shows its location within the structural framework of Hercynian mobile belts¹. Further details on shallow and deep data available at the start of this work will be given elsewhere².

The field operations and processing were carried out between late 1983 and early 1985. The Compagnie Générale de Géophysique (CGG) did most of the seismic work, the Institut National des Sciences de l'Univers (INSU) did the complementary wide-angle reflection and refraction survey, and the Laboratoire de Géophysique Appliquée et Structurale de Nancy (LAGAS) did the magnetotelluric survey. The same layout (192 groups, 80 m between groups) was used by CGG for each type of seismic recording (reflection, refraction and wide-angle). For the reflection survey, five heavy vibrators, each weighing 13.5 tonnes and spaced 8 m apart, produced a 20-s sweep. One shotpoint was made by stacking 10×5 vibrations, providing a 96%-fold CDP (common-depth-point) record. As a complement, direct and inverse refraction lines were shot at 15-km intervals, using 50–300 kg of dynamite. The wide-angle data were obtained using a source consisting of 250–600 kg of explosives located ~100 km away from the recording equipment, to be at the critical distance for the Moho depth. The magnetotelluric data were recorded at 15 stations set out every 15 km along the seismic profile. Records were taken along the strike of the line (N45° E), and orthogonally (N135° E), to test for the existence of any anisotropy. Full details of geoelectric studies will be published elsewhere. We describe here the best-fitting two-dimensional model of the magnetotelluric response. Published gravimetric and magnetic maps of France (1:1,000,000 scale) were used to compute the possible geometry of bodies at the origin of anomalies. The main constraint was the shape of the seismic reflectors observed, although an exact depth correlation was not possible.

Figure 2 shows the main results and geophysical interpretations. Due to the crustal heterogeneities observed, interval velocities have had to be smoothed. As an example, an average velocity of 3.3 km s⁻¹ for the Mesozoic section is known from stacking velocities of commercial lines. Beneath the Mesozoic, velocities were obtained from wide-angle reflection. A 5.5 km s⁻¹ velocity is proposed for the sedimentary or slightly metamorphosed Palaeozoic, a 6.0 km s⁻¹ velocity in the upper crust, a 6.7 km s⁻¹ velocity in the lower crust down to the Moho, and a 6.2 km s⁻¹ velocity for the crust as a whole. For the sake of the discussion, in Fig. 2, four geological units are delineated:

(1) Down to 1 s, the Mesozoic and Cenozoic cover of the Paris Basin displays the main sedimentary layers calibrated by oil wells, as well as the main surface faults such as the Seine, Bray and Somme faults. These controlled Mesozoic sedimentation as normal faults, and had reverse or transcurrent displacement during the Eocene 'Pyrenean' shortenings. The vertical displacements, however, remain quite small: a maximum offset of only 200 ms has been detected on the Bray fault. That these faults are probably late Hercynian features has been confirmed by the study of deep samples from a borehole close to the Bray fault. Precambrian granites (570 Myr) were subsequently transformed into a mylonitic orthogneiss in Hercynian times (330 Myr). The syn-metamorphic microfabrics are related to dextral strike-slip motion³.

(2) At the northern end of the profile (shotpoints (SP) 100



Fig. 1 Structural map of the Hercynian fold belt of western Europe (after ref. 3) and location of the ECORS Nord de la France deep seismic profile. 1, Major thrusts; 2, inner crystalline nappes and sutures; 3, areas affected by folding and schistosity; 4, outer Carboniferous basins; 5, platform and central zone, undeformed or slightly deformed during the Variscan orogeny; 6, direction of major overthrusts; 7, ductile shear zone; 8, Nord de la France profile.

to 1,500) and down to 3 s, a seismically transparent seismic sequence is cut by some southward-dipping reflectors. This sequence is interpreted as the allochthonous Palaeozoic Dinant Nappe, which, until now, has been only roughly suggested by a few oil wells and seismic reflection profiles shot further north by industry. Its top part outcrops along the Meuse River, where late tectonic deformation has been dated as late Westphalian (Asturian phase). The nose of the nappe is considered to represent the Variscan front outlined by the Midi fault. This is a flat overthrust correlating on the line drawing to the top of a reflective sequence ~1 s thick, outcropping near SP 50 and continuing at depth up to SP 1,500 (that is, to >90 km). This seismic sequence is interpreted as the autochthonous or little-deformed Upper Palaeozoic sediments preserved in place below the nappe which were encountered at the Jeumont borehole a few tens of kilometres east of the line. These sequences are also conspicuous in the magnetotelluric data, where they are correlated with a highly conductive zone (resistivity = 30 Ω m) ~3–6 km thick. Below 5 s, another seismically transparent zone corresponds to the Caledonian Brabant block, where folding and schistosity developed in Silurian times.

(3) In the central part of the profile (SP 1,500 to 3,300), the overthrusting belt shows some remarkable southward-dipping reflectors between 3 and 8 s. In the absence of any geological check, they are interpreted as a set of large imbricated thrusts representing the deep roots of the Dinant Nappe. The modelling of gravimetric and magnetic anomalies can help in the interpretation in this area. The external shape of the bodies at the origin of these anomalies can be contoured as overthrusting slabs. Some of them could be regarded as ophiolitic sheets, such as that described⁴ further west in the Lizard complex in Cornwall (southwestern United Kingdom). Magnetotelluric results show that this area appears to be more conductive and heterogeneous than the surrounding one.

From 8 to 12 s, the stratified lower crust clearly contrasts with the transparent one to the north (Brabant block). The origin of

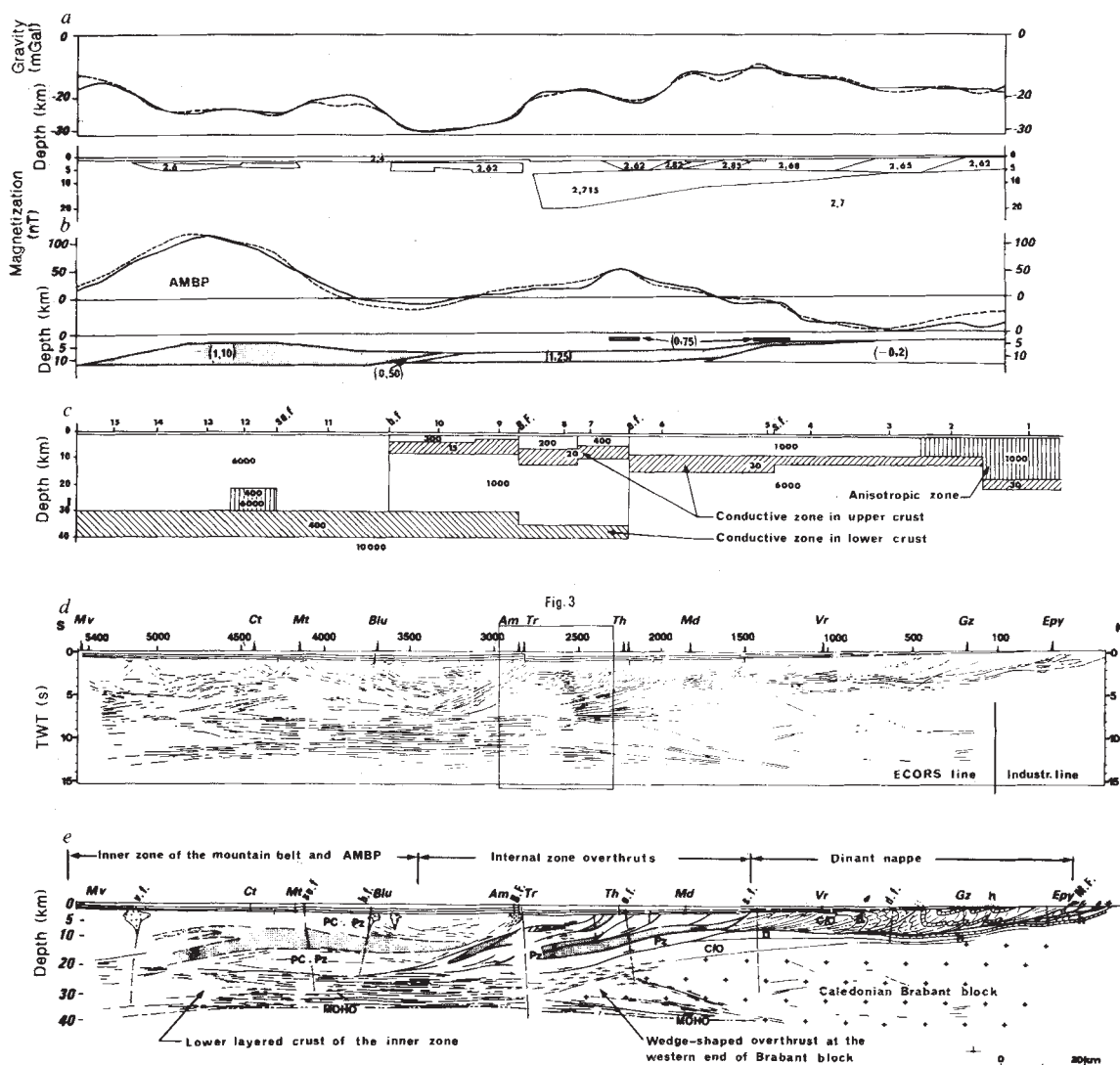


Fig. 2 Interpretation of the ECORS Nord de la France profile. *a*, Gravimetric profile and modelling; numbers are specific gravities (reference s.g. = 2.7). *b*, Aeromagnetic profile (altitude, 3,000 m) and modelling; numbers are magnetization in $A m^{-1}$. AMBP, anomalie magnétique du Bassin de Paris. *c*, Geoelectrical section along the profile from magnetotelluric two-dimensional modelling (resistivities in Ωm). *d*, Two-way time line drawing (unmigrated). On top are the common-depth-point numbers and boreholes. *e*, Geological interpretation. h, Carboniferous; d, Devonian; C/O, Cambro-Ordovician; PZ, metamorphic undifferentiated Palaeozoic; PC-PZ, undifferentiated Precambrian and Lower Palaeozoic of the inner metamorphic zone; grey pattern, magnetic rocks (banded iron formation or remnant oceanic crust); M.F., Midi Fault; B.F., Bray Fault; d.f., Doullens Faults; s.f., Somme Fault; e.f., Eu Fault; b.f., Banthelu Fault; se.f., Seine Fault; v.f., North Marville Fault.

this layering is still subject to debate. Whereas many authors have argued that it could be related to the Hercynian orogeny (tectonic lamination, mantle intrusions, and so on), we suggest that it could be of recent age and the result of either some upward 'Moho readjustment' after the Hercynian crustal stacking or lower-crustal ductile flow during Permian or Mesozoic sedimentary basin subsidence. The lower boundary of the stratified crust defines the Moho at an average travel time of 12 s (~ 37 km). Its apparent slight northward dip could possibly be due to a difference in velocities between two different crusts. Lastly, at SP 2,820, a transparent zone in the lower crust appears below the Bray fault. It was initially interpreted as a shear zone related directly to the prolongation of the fault at depth². However, wide-angle reflection data have shown that this transparent zone is reduced in width and that no vertical displacement of reflectors can be discerned. The trace of this fault in the lower crust remains debatable.

(4) At the southern end of the profile (SP 3,300 to 5,441) and from a depth of 1 to 8 s, the presence of reflectors with opposite dips is more difficult to interpret. Metamorphic rocks found by drilling reveal that we are in an inner part of the chain. Some

seismic features can be interpreted as granitic intrusions, examples of which have been found at the bottom of boreholes. These could be the cause of some well-known negative gravimetric anomalies. The remaining major problem is the interpretation of the large 'anomalie magnétique du Bassin de Paris' (AMBP). The modelling of the magnetic body as a tabular element corresponding to the transparent area between 2.5 and 5 s can tentatively be ascribed to an overthrust belt of Palaeozoic basic rocks⁵. Alternatively, it could be related to Precambrian magnetic sedimentary rocks, such as a banded iron formation. The stratified horizons below 8 s are part of the lower crust. They correspond in the magnetotelluric data to a conductive sequence near the bottom of the crust, as observed in other tectonically active regions^{6,7}.

Our data allow significant comparisons to be made with similar sections recently acquired in the United Kingdom^{8,9} and West Germany¹⁰, across the Variscan chain, as well as in the Appalachian orogen on the other side of the Atlantic Ocean. A better understanding of the geological history of the Variscan belt and its sedimentary cover in northern France is now possible. The Hercynian mountain belt extends for 3,000 km, from

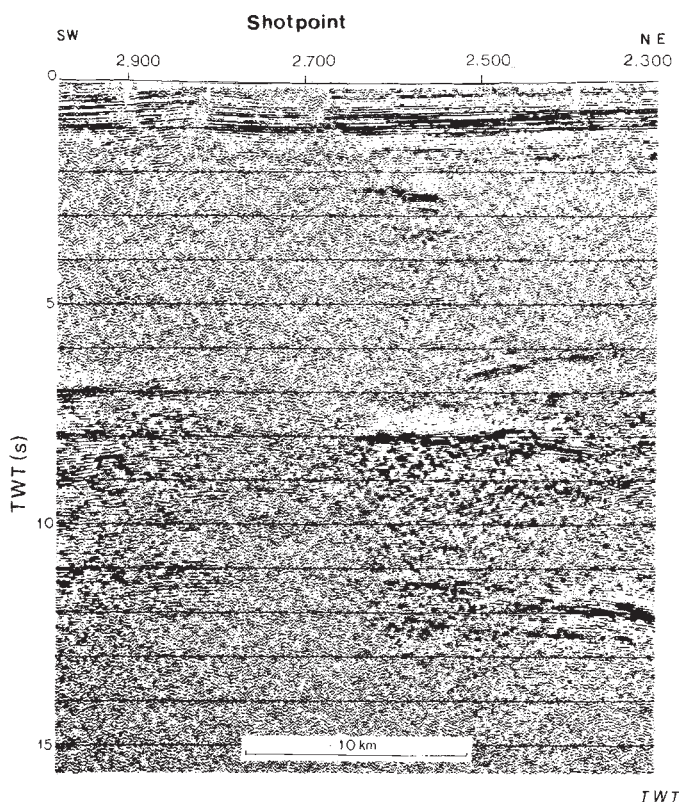


Fig. 3 Deep seismic section across the Bray fault (area outlined by box in Fig. 2d). On the left-hand side: 0–1 s, Mesozoic cover; 1–7 s, upper crust; 7–12 s, lower crust. The Moho deepens slightly from 12 to 12.5 s from south-west to north-east. The Bray fault is at SP 2,800.

the southern end of the Iberian Peninsula to the Bohemian Massif in Czechoslovakia (Fig. 1). Two mobile areas on either side of a central Cadomian platform show southward and northward (Variscan) overthrusting, where basic and high-pressure metamorphic rocks are encountered. The Nord de la France profile crosses the northern segment, which is supposed to have originated when a small oceanic basin opened in middle (?) Devonian times. A remnant of its crust can be found in the 375 ± 34 -Myr-old Lizard ophiolites in south-west Cornwall, or in the ophiolitic klippen in the Münchberg Massif (Germany). This oceanic domain had closed by the end of the Devonian, by subduction and collision, the chronology of which has recently been improved³. The related suture zone could be depicted in our line by the southward-dipping reflectors south of the Bray fault in the lower crust and in the upper mantle. From early Dinantian to late Westphalian, the collision was followed by continental shortening which led to large overthrust, decollement and thin-skin tectonics over most of the area between the Bray fault and the Brabant Massif foreland. Such a tectonic style has also been described to the east in West Germany^{10–12}, whereas, to the west, the SWAT profiles in the Celtic Sea seem to depict a more rapid deepening of major thrust zones⁹ in the crust. The chain was probably deeply eroded in the Stephanian and Permian. Except for some isolated Permian troughs only a few hundred metres thick, subsidence started in the Triassic on what was to become the Paris basin. Sedimentation was locally controlled by normal faults (Seine, Bray, Somme faults), which were reactivated in the Eocene as transcurrent or reverse faults. This led to a slight local inversion of the basin between the Seine River and the Pays de Bray.

The ECORS project is being carried out by the association of the Institut Français du Pétrole (IFP), the Institut National des Sciences de l'Univers (INSU-CNRS), the Société Nationale Elf-Aquitaine (ELF), and the Institut Français pour l'Exploitation de la Mer (IFREMER).

Received 24 January; accepted 29 May 1986.

1. Matte, Ph. *Am. geophys. Un. Geodyn. Ser.* 10, 53–81 (1983).
2. Bois, Ch. *et al. Am. geophys. Un. Geodyn. Ser.* Vol. 13 (1986).
3. Matte, Ph. *et al. Bull. Soc. géol. Fr.*, Ser. 8, vol. II (1986).
4. Le Gall, B. thesis, Univ. Brest (1983).
5. Castaing, C. *Docum. BRGM No. 81-2*, 163–184 (1984).
6. Pham, V. N. *et al. C.r. hebdo. Seanc. Acad. Sci., Paris* 287, 199–202 (1978).
7. Pham, V. N. *Earth Planetary Sci. Letters*, 52, 373–380 (1981).
8. Whittaker *et al. Bull. Soc. géol. Fr.*, Ser. 8, vol. II (1986).
9. BJRPS & ECORS *J. geol. Soc.* 143 (in the press).
10. DEKORP Research group *J. Geophys.* (in the press).
11. Nelson, K. D. *et al. Geology* 13, 714–718 (1985).
12. Meissner, R. & Wever, Th. *Am. geophys. Un. Geodyn. Ser.* vol. 13 (1986).

Collapse of the Caledonian orogen and the Old Red Sandstone

K. R. McClay*, M. G. Norton*, P. Coney†
& G. H. Davis†

* Department of Geology, University of London Royal Holloway and Bedford New College, Egham TW20 0EX, UK

† Department of Geological Sciences, University of Arizona, Tucson, Arizona 85721, USA

The classic Caledonian red sandstones, breccias and conglomerates of the Devonian 'Old Red Sandstone' sequences of north-west Scotland, Norway and East Greenland have long been regarded as traditional 'molasse'-type deposits generated by the uplift of the Caledonian mountain chain^{1–4} and controlled either by extensional^{5,6} or strike-slip^{7–9} faulting. Here we propose that the Old Red Sandstone was deposited in extensional basins developed as a result of the collapse of the overthickened crustal welt which resulted from Caledonian compressional tectonics. The Devonian basins developed within the Caledonian mountain chain very soon after the cessation of compressional tectonics, locally reactivating these compressional structures as extensional fault systems. We review the tectonic setting of the Old Red Sandstone in the Caledonian orogen and suggest that the most applicable tectonic analogue is the intra-orogen collapse and extension model proposed for the mid-Tertiary Basin and Range province in the western United States¹⁰. This tectonic model has important implications for subsequent Mesozoic extension during the evolution of the North Sea basin.

The term 'Old Red Sandstone' (ORS) has been used since 1822¹¹ to describe the Devonian-age, largely terrestrial, red-bed facies of sandstones, conglomerates, breccias and volcanic rocks of northern Britain⁵. Similar Devonian terrestrial red-bed facies are found in East Greenland¹² and Norway^{13–15}. The distribution of Devonian continental sediments has been used to define an 'Old Red Sandstone continent' that extended from north Greenland through Norway, north-west Scotland, Ireland and eastern Canada, and southwards into the Appalachians². Here we restrict our discussion to the northern part of the Caledonides and the Devonian sedimentation developed on the Caledonian orogen infrastructure, where Devonian sediments are found in fault-bounded basins within the orogen (Fig. 1).

In north-west Scotland, ORS deposition occurred in a number of fault-bounded basins, of which the Orcadian Basin and the Midland Valley graben are the most important (Fig. 1), having received >9 km of terrestrial sediments⁴. The ORS has been subdivided into the Lower, Middle and Upper divisions (see ref. 4) and not all divisions are developed in every basin^{4,9}. The Lower and Middle ORS are separated by folding and faulting^{4,9,16}, which produced unconformities and non-deposition within the ORS sequence. Post-Middle-ORS tectonism appears to be minor⁴. In addition to the major ORS basins, many smaller fault-bounded basins are found within the north-west Scottish highlands^{4,16} and also off-shore (Fig. 1), where seismic reflection profiling has indicated that the ORS was deposited in half-

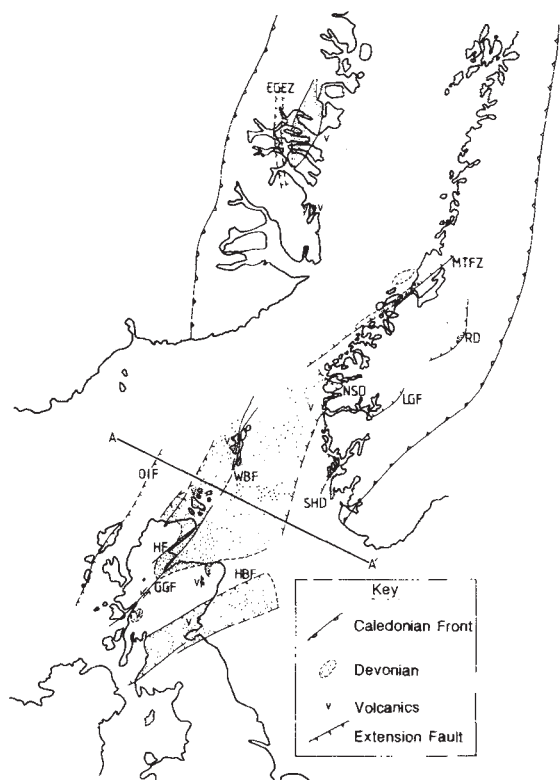


Fig. 1 Tectonic map of the central and northern Caledonides, showing the major zone of Caledonian compressional tectonics. Approximately 100 km (conservative estimate) of Mesozoic extension has been removed, bringing Norway closer to north-west Britain and East Greenland. Devonian ORS extensional basins and Devonian volcanics are shown. Major structures that were probably active in the Devonian are marked: HBF, Highland Boundary Fault; GGF, Great Glen Fault; HF, Helmsdale Fault; WBF, Walls Boundary Fault; OIF, Outer Isles Fault (commonly termed the Outer Isles 'Thrust'); SHD, Sunhordaland Detachment (affecting Late Silurian granites); LGF, Laerdal-Gjende Fault; SD, Sunnfjord Detachment²¹; MTFZ, More-Trondelag fault zone¹⁹; EGEZ, East Greenland extensional zone^{12,23,24}.

grabens which formed by reactivation of Caledonian thrust faults¹⁷. In all cases, thick wedges of fault-controlled terrestrial sediments are found in basins that generally trend NNE, parallel to the Caledonian structural grain. Calc-alkaline volcanics associated with Lower ORS deposition are abundant in the Midland Valley and in Shetland (see refs 4, 9, 16). Watson^{9,16} has postulated that considerable uplift occurred in the Caledonian orogen in the period 500–420 Myr and that the present erosion surface of the Highland massif corresponds closely to the Devonian land surface. Thus, ORS deposition in northern Britain was characterized by terrestrial sedimentation in fault-bounded basins within the Caledonian mountain belt. In some cases, reactivation of Caledonian structures controlled the Devonian extensional faults. It has recently been proposed that the Outer Isles fault zone is a major extensional structure of Mesozoic age, as indicated by inferred Moho displacement on the MOIST seismic line¹⁷. Phyllonites from the fault zone show an extensional sense of movement and give K–Ar ages of 395–410 Myr (ref. 18), suggesting an earlier phase of extensional activity during the Lower Devonian.

In Norway, non-marine clastic ORS sediments occur in small fault-bounded basins within the Caledonian orogen in the Nordfjord-Sognefjord, Trondheim and Roragen areas¹⁹. Their sedimentary fill ranges from fanglomerates to sandstones, and varies in thickness from 1 to ~5 km, although original thicknesses may have ranged up to 10 km (ref. 19). The southernmost basin (Solund) contains volcanic units interbedded with the marginal fanglomerates²⁰. The most intensely studied of these basins, the Hornelen basin^{14,15,19}, contains spectacular examples of fault-controlled marginal alluvial fans. Steel¹⁹ argues that deposition in the Hornelen basin was strongly controlled by dextral strike-slip faulting; however, reappraisals of the tectonics of these basins (refs 21, 22) have clearly demonstrated that extensional tectonics dominate and that the basin-margin east-west faults are most likely to be lateral ramps and link/transfer faults in a larger extensional fault system. In particular, Hossack²¹ emphasizes the listric growth-fault geometry in these basin fault systems. This research has led to the identification

of a number of Devonian extensional structures throughout the southern Scandinavian Caledonides (Fig. 1).

The Caledonides of East Greenland display spectacular examples of ORS deposition and tectonics^{12,33}. At least 7 km of breccias, conglomerates and sandstones are developed in the Devonian basin in central East Greenland. Haller¹² describes this succession as internal 'molasse', deposited in a series of graben basins within the Caledonian orogen. Extensional faulting abounds^{12,23,24}, both within the basin and in the Caledonian basement underlying it. Two sets of extensional faults have been documented. An early low-angle set of extensional faults strikes NNW–SSE and is cut by a later set of high-angle extensional faults that strike N–S or NNE–SSW^{12,24}. These faults exhibit typical 'Basin and Range' geometries^{25,26}, with linked listric extensional faults cut by later high-angle planar faults (Fig. 2). Frankl²⁴ calculated a total extension of 42 km on these fault systems, with kilometric displacements on individual faults. In addition to the tectonic features of this major zone of crustal extension in East Greenland, there is associated Devonian volcanism with calc-alkaline, alkaline and rhyolitic volcanics¹².

Thus there is significant evidence for major crustal extension during the deposition of the ORS, which generated basins within the Caledonian orogenic belt. In addition to the extensional character of the Devonian there was significant strike-slip fault-

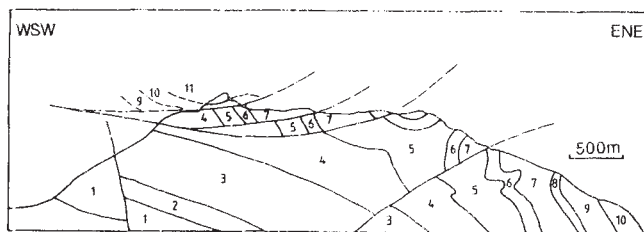


Fig. 2 Sketch from ref. 24 showing a listric fault fan extending Upper Eleonora Bay Group sediments in northern Scoresby Land, East Greenland. Units are marked 1 to 11 to indicate relative stratigraphic positions.

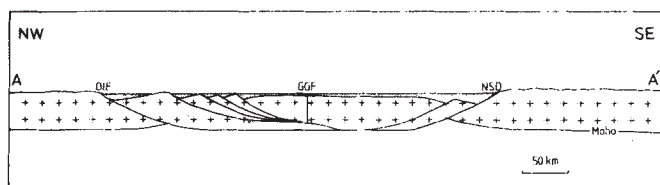


Fig. 3 Sketch cross-section through the British and Norwegian Caledonides, before Mesozoic extension, showing deep-seated Devonian detachment faults, generated by the collapse of over-thickened Caledonian crust. No vertical exaggeration.

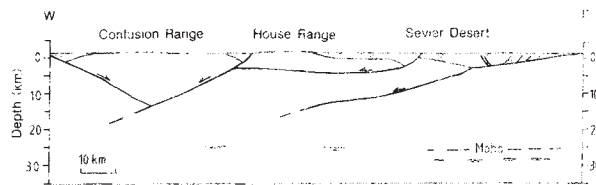


Fig. 4 Idealized sketch of part of the Basin and Range province, adapted from ref. 39. Stippled areas, mid-Tertiary and younger basin fill; blank area, underlying Palaeozoic strata and basement. The Moho occurs at ~25 km depth³⁹.

ing, for example, on the Great Glen Fault in northern Scotland, the Walls Boundary Fault in the Shetlands and also in the Greenland Caledonides¹². We suggest that these strike-slip faults, which run parallel to the trend of the extensional faults, formed separate and, in terms of basin formation, subordinate tectonic features within the extending orogen.

Appraisal of the Tertiary extensional tectonics of the southwestern United States, superimposed on an earlier Mesozoic-early Tertiary contractional tectonic phase^{10,27,28}, suggests a direct comparison with the Devonian ORS tectonics in the central and northern Caledonian orogen. The geometries of extensional fault systems associated with Cordilleran metamorphic 'core complexes' and the Basin and Range province have been well documented^{25,26,29,30}, and similar styles can be demonstrated for the ORS tectonics (Fig. 2). The Tertiary to Recent desert gravels, fanglomerates and playa lake sequences in the extensional province of the western United States can be directly compared to the terrestrial red-bed sequences that characterize the ORS. Similarly, the western-American extensional province is characterized by abundant volcanism interbedded with syntectonic basin fill³¹, a direct analogue of the Devonian volcanism in Scotland³², East Greenland¹² and western Norway²⁰.

Coney²⁷ and Coney and Harms¹⁰ have proposed that the middle Tertiary metamorphic core complexes and late Tertiary Basin and Range extension in the southwestern United States are a direct result of the collapse of an overthickened crustal welt which was generated by the Mesozoic-early Tertiary Laramide compressional tectonics. They envisage a crustal welt 50–60 km thick which became the site of deep-seated Tertiary extension as the result of lateral spreading and gravitational instability¹⁰. A similar model may be erected for the Devonian of the central and northern Caledonides. In the Scottish Caledonides, closure of the Iapetus ocean and crustal thickening by telescoping of Lewisian basement slices and Moine and Dalradian metasediments over the Lewisian foreland resulted in a thickened crustal welt by end-Silurian times (420 Myr)^{9,33}. In the Norwegian Caledonides, ages from eclogites formed during the extreme crustal thickening date the end of the main compressional phase as 425 Myr (ref. 34). This overthickened welt would tend to spread laterally if there were sufficient lateral density variation, topographic head and lowering of viscosity^{10,35}. Recent work on the strength of thickened lithosphere³⁶ suggests that significant weakening will occur 15–20 Myr after the main thrusting. This is in general agreement with the relative timing of compression and extension in the Caledonides. Thus, we can envisage pure shear deformation of the lower crust, with a surface expression of listric extensional faulting and basin development as the overthickened crust collapses to a more stable thickness of 30–40 km. In the southwestern United States, lower-crustal melting accompanied this spreading process, as evidenced by the late Mesozoic-Tertiary plutons^{10,27}. Similarly, the late Caledonian to Devonian plutonism also attests to crustal melting³⁷. The analogies between the tectonic evolution of the ORS and the Tertiary extensional province of the southwestern United States are thus very close. The Devonian in the Caledonides was a period of deep-seated crustal extension

that restored the crust to a more normal thickness before the later Mesozoic crustal thinning which formed the North Sea basin^{5,9}. The Midland Valley basin which formed at the southern margin of the thickened Caledonide crustal welt falls within the same overall extensional setting. Its association with 'Cascade'-type magmatism perhaps suggests that its formation is more closely related to extension caused by continued but waning subduction after collision³⁸.

Figure 3 shows a hypothetical crustal section through the British and Norwegian Caledonides, which may be compared with a simplified sketch, based on seismic reflection profiling, of a crustal section from part of the Basin and Range province (Fig. 4). In the Basin and Range, major extensional faults are thought to extend to the lower crust³⁹, and we postulate a similar geometry for the Devonian extensional structures in the Caledonides (Fig. 3). The Devonian structures of the Caledonian orogen (Figs 1, 3) are the architecture on which the later Mesozoic extension of the North Sea basin developed, and are thus likely to have had significant effects on the geometry of basin development in the North Sea.

Received 1 July 1985; accepted 30 May 1986.

1. Leeder, M. R. *J. geol. Soc. Lond.* **139**, 481–494 (1982).
2. Friend, P. F. *Mem. Am. Ass. Petrol. Geol.* **12**, 703–710 (1969).
3. Harris, A. L. *Mem. geol. Soc. Lond.* **9** (1985).
4. Mykura, W. in *Geology of Scotland* 2nd edn (ed. Craig, G. Y.) 205–251 (Scottish Academic, Edinburgh, 1983).
5. Dewey, J. F. *J. geol. Soc. Lond.* **139**, 371–414 (1982).
6. Watson, J. V. & Plant, J. *Phil. Trans. R. Soc. A* **291**, 321–338 (1979).
7. Bluck, B. J. *Spec. Publ. int. Ass. Sedimentol.* **4**, 63–78 (1980).
8. Zeigler, P. A. *Geological Atlas of Western and Central Europe* (Elsevier, Amsterdam, 1982).
9. Watson, J. V. *J. geol. Soc. Lond.* **141**, 193–214 (1984).
10. Coney, P. J. & Harms, T. A. *Geology* **12**, 550–554 (1984).
11. Conybeare, W. D. & Phillips, W. *Outlines of the Geology of England and Wales* (London, 1822).
12. Haller, J. *Geology of the East Greenland Caledonides* (Wiley, London, 1971).
13. Nilsen, T. H. *Norg. geol. Unders.* **259**, 1–108 (1968).
14. Steel, R. J. *Tectonophysics* **36**, 207–224 (1976).
15. Steel, R. J. & Gloppen, T. G. *Spec. Publ. int. Ass. Sedimentol.* **4**, 79–103 (1980).
16. Watson, J. V. *J. geol. Soc. Lond.* **142**, 221–244 (1985).
17. Brewer, J. A. & Smythe, D. K. *J. geol. Soc. Lond.* **141**, 105–120 (1984).
18. Sibson, R. H. *J. geol. Soc. Lond.* **131**, 183–201 (1975).
19. Steel, R. J. *Geol. Surv. Pap. Can.* **78-13**, 57–62 (1978).
20. Sigmond, E. M. O., Gustavson, M. & Roberts, D. *Berggrunnskart over Norge*, 1:1 million (Norges Geologiske Undersokelse, Trondheim, 1984).
21. Hossack, J. R. *J. geol. Soc. Lond.* **141**, 629–638 (1984).
22. Norton, M. G. *Tectonics* **5**, 195–204 (1986).
23. Wegmann, C. E. *Meddr Grønland* **103**, 1–148 (1935).
24. Frankl, E. *Meddr Grønland* **113**, 1–160 (1953).
25. Proffett, J. M. *Geol. Soc. Am. Bull.* **88**, 247–266 (1977).
26. Wernicke, B. & Burchfiel, B. C. *J. struct. Geol.* **4**, 104–115 (1982).
27. Coney, P. J. *Continental Extension Tectonics Meet.*, Durham, abstr. (Geological Society, London, in the press).
28. Davis, G. H. & Coney, P. J. *Geology* **7**, 120–124 (1979).
29. Anderson, R. E. *Geol. Soc. Am. Bull.* **82**, 43–58 (1971).
30. Dickinson, W. R. *Geol. Soc. Am. Abstr. Progr.* **16**, 488 (1984).
31. Gans, P., Mahood, G. & Schermer, E. *Continental Extension Tectonics Meet.*, Durham abstr. (Geological Society, London, in the press).
32. Thirwall, M. F. *J. geol. Soc. Lond.* **138**, 123–138 (1981).
33. Powell, D. & Phillips, W. E. A. *Mem. geol. Soc. Lond.* **9**, 17–40 (1985).
34. Griffin, W. L. et al. in *The Caledonide Orogen: Scandinavia and related areas* (eds Sturt, B. A. & Gee, D. G.) (Wiley, New York, 1985).
35. England, P. in *Mountain Building Processes* (ed. Hsu, K. J.) (Academic, New York, 1982).
36. Glazner, A. F. & Bartley, J. M. *Geology* **13**, 42–45 (1985).
37. Brown, G. C., Francis, E. H., Kennan, P. & Stillman, C. J. *Mem. geol. Soc. Lond.* **9**, 1–16 (1985).
38. Zhou, Jiaxin *J. geol. Soc. Lond.* **142**, 309–318 (1985).
39. Allmendinger, R. W. et al. *Geology* **11**, 432–436 (1983).

Palaeoclimatic changes deduced from $^{13}\text{C}/^{12}\text{C}$ and C/N ratios of Karewa lake sediments, India

R. V. Krishnamurthy, S. K. Bhattacharya & Sheela Kusumgar

Physical Research Laboratory, Ahmedabad 380 009, India

Most long palaeoclimatic records based on isotope studies have come from deep-sea cores; however, in order to obtain information on regional climatic variations, it is essential to study continental deposits. The chief limitation has been the lack of well-preserved datable geological records, with the exception of cave formations. We report here values for the stable carbon isotope ratio, $^{13}\text{C}/^{12}\text{C}$, and the elemental carbon to nitrogen (C/N) ratio of the organic fraction of the palaeomagnetically and radiometrically dated Pleistocene Karewa lake sediments from Kashmir, India. Together, the two ratios, which are anticorrelated, provide an excellent marker of the source of organic matter in the lake. The Karewa sediments indicate at least three major episodes of shallower than normal levels. These episodes of shallow lake levels, and hence a cold, dry climatic regime, are broadly correlatable with deep-sea records.

Samples for this study were collected from the Romushi section, one of several tectonically exposed localities in the intermontane Kashmir basin in northwestern India. The boat-shaped Kashmir basin ($33^{\circ}30' - 30^{\circ}30' \text{ N}$, $74^{\circ} - 75^{\circ}30' \text{ E}$) comprises unconsolidated gravel-sand-mud successions which have come to be collectively known as the Karewa sediments, the term Karewa meaning plateau in the local language. The basin once formed a vast lake, $\sim 5,000 \text{ km}^2$ in area, until a breach by the modern Jhelum river caused it to be drained. Later tectonic events have exposed these sediments at various localities in the valley. The total thickness of such sediments may exceed 1 km. The Karewa sediments have generated considerable interest for more than a century, largely due to the presence of mammalian fossils and the possibility of deciphering Pleistocene glacial history in the region¹. Lithologically, the Karewas have been divided into the Lower and Upper Karewas, with a loess layer marking the top of the formation². Controversy exists regarding the depositional environment of the Karewas during different periods. The Romushi section, which exposes the Lower Karewas, is generally regarded to be predominantly lacustrine in origin³. Recently, radiometric and palaeomagnetic dating of the Karewa sediments^{4,5} has yielded an age of $\sim 4 \text{ Myr}$.

Pollen studies of various sections in the Kashmir basin suggest that the palaeoclimatic pattern there follows a global trend. For example, the Pliocene warming and the glacial/interglacial oscillations of the Pleistocene could be discerned⁶. Previous studies using stable carbon isotope ratios of palaeosols⁷ and pollen from peat bogs⁸ point to an early end of the last glaciation. Thus, the Karewas provide a unique opportunity to investigate Plio-Pleistocene climatic changes in the valley.

The sediment characteristics at Romushi have been dealt with in detail previously^{3,9,10}. Fourier analysis¹¹ of grain size distributions for subsections where sedimentation rates were apparently constant, seems to indicate that cycles with Milankovitch orbital periodicities may be present in the sedimentary record.

The $^{13}\text{C}/^{12}\text{C}$ and C/N ratios were measured on the samples that were used for magnetic measurements, thereby restricting isotope and elemental analysis to fine-grained mudstone and silt facies. This selection of one sediment type, however, assures uniform sample characteristics. Magnetostratigraphy and fission track dating of volcanic ash embedded in the sedimentary column have been discussed elsewhere^{12,13}. A mean fission track age of $2.3 \pm 0.3 \text{ Myr}$ obtained by us for the two ash layers

compares well with an earlier estimate of $2.4 \pm 0.2 \text{ Myr}$ (ref. 5). For isotope measurements of organic carbon, samples were first demineralized with 20% HCl and dried. A portion of the sample was combusted to give CO_2 gas for mass spectrometric measurements; the C/N ratios were determined on another portion by a micro-Kjeldhal technique.

The $\delta^{13}\text{C}$ values and C/N ratios are plotted as a function of depth in Fig. 1. $\delta^{13}\text{C}$ is defined as

$$[(^{13}\text{C}/^{12}\text{C})_{\text{sample}} / (^{13}\text{C}/^{12}\text{C})_{\text{PDB}} - 1] \times 10^3$$

where PDB is the international limestone standard for comparing stable carbon isotope ratios. The overall precision of $\delta^{13}\text{C}$ measurements for field samples is $\pm 0.2\%$. Also shown at the top of Fig. 1 is the magnetostratigraphy of the section, indicating the age of the ash and other known ages of various boundaries¹⁴. Additional details pertaining to these aspects are discussed in ref. 4. The $\delta^{13}\text{C}$ values range from -19 to -27% and the C/N ratios from 0.33 to 20, and the two are anticorrelated, with a correlation coefficient of -0.56 (significant at the 2% level).

The lacustrine environment is a very complicated hydrological and biological system, which is reflected in the isotopic composition of its sediments, especially the organic fraction. The main sources of organic carbon in a lake are submerged plants, such as macrophytes, floating pond weeds or planktonic material, and terrestrial vegetation¹⁵. Of these, the submerged plants tend to be isotopically heavier because their source of CO_2 is dissolved bicarbonate rather than atmospheric CO_2 . The enrichment of ^{13}C in the process $\text{CO}_2 \rightarrow \text{HCO}_3^-$, the latter being heavier, results in the fixed carbon also being enriched. Similarly, the $\delta^{13}\text{C}$ value of planktonic material reflects the plankton's source of carbon: use of dissolved bicarbonate results in ^{13}C enrichment; use of atmospheric CO_2 leads to an overlap of planktonic $\delta^{13}\text{C}$ with the values for terrestrial plants. The latter fall into the well-known C_3 and C_4 categories, with characteristic average $\delta^{13}\text{C}$ values of -27 and -14% , respectively^{16,17}. A temperature-dependent enrichment may be present, as well as opposing effects caused by fermentation of lake mud and methanogenesis, but their full role is not yet clear. A recent survey¹⁸ of the $\delta^{13}\text{C}$ values of organic matter from eight southern Swedish lakes considers many of these factors, and there is a general indication that most parameters that affect the isotopic distribution of lacustrine organic matter remain relatively constant, changing only if the environmental conditions are altered to a significant degree¹⁹.

Nakai's early work²⁰ on the 200-m-thick lake Biwa sediments was one of the successful applications of carbon isotopes in lake organic matter to palaeoclimatology. He observed a good correlation between the $\delta^{13}\text{C}$ values of the organic fraction and the absolute carbon content, which could be explained in the following way. During warmer periods the lake productivity was high, resulting in an increased carbon content of total sediment and more dissolved bicarbonate contributed to the lake's carbon pool to keep pace with the requirements of a highly productive lake. During a cold phase the opposite effect operated, making the organic carbon relatively depleted in ^{13}C and also causing a decrease in the absolute carbon content. Later extensive work by Stuiver¹⁵ showed the above model to be true in only a few cases. Nevertheless, this and more recent work indicated a definite depletion in ^{13}C content at times of cooler climate^{19,21}.

In a refined approach, Nakai and co-workers^{22,23} suggested the use of the C/N ratio in combination with the $\delta^{13}\text{C}$ value of lake organic matter, to interpret the latter in terms of palaeoenvironmental changes. Lower organisms such as algae, phytoplankton and zooplankton are characterized by a high protein content, hence low C/N ratios, while terrestrial plants, which are the main source of terrestrial organic matter, are characterized by a low protein content, hence high C/N ratios; thus, the C/N ratios of bottom sediments can be used as a powerful complement to their $\delta^{13}\text{C}$ values in evaluating the relative

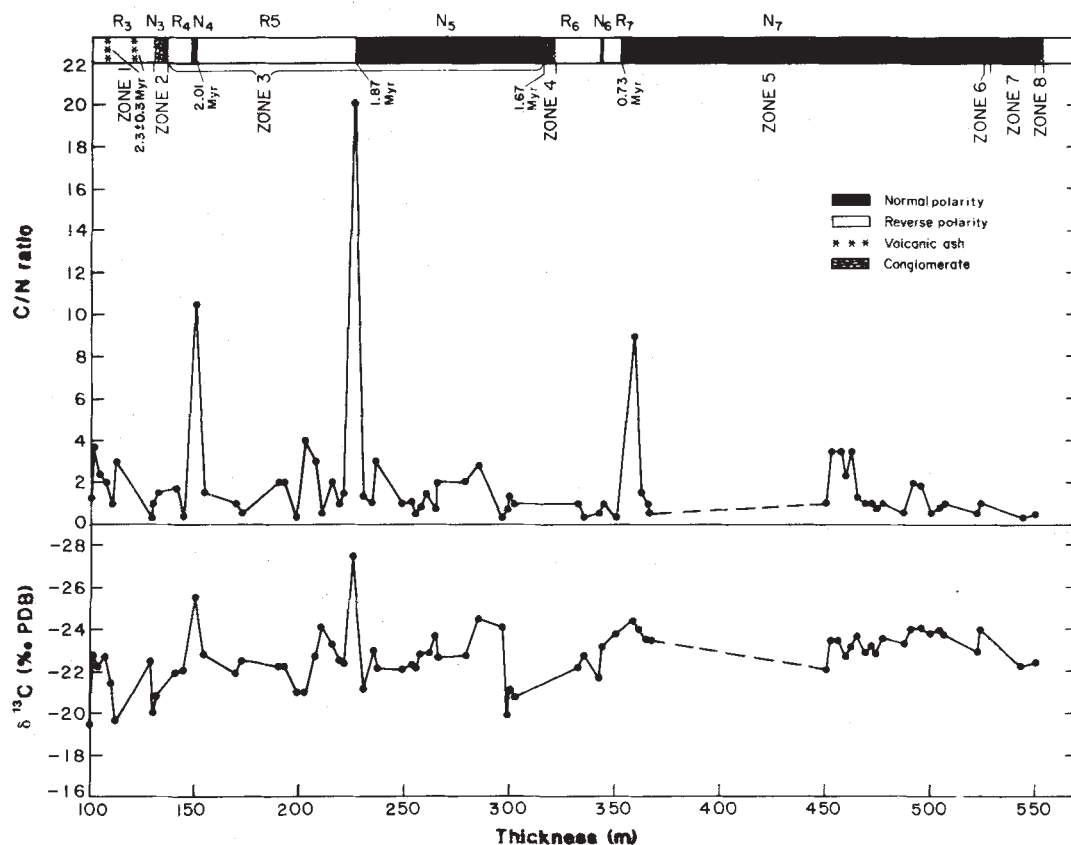


Fig. 1 C/N ratio and $\delta^{13}\text{C}$ of the organic fraction of Karewa lake sediments of the Romushi section. The magnetostratigraphy of the same section is shown at the top. The section represents a time interval from 2.4 Myr at the base to ~300–400 kyr at the top (zone 8), the latter being a conservative estimate based on extrapolation of a highly variable sedimentation rate¹⁰.

proportions of autochthonous and allochthonous carbon sources.

The $\delta^{13}\text{C}$ values and C/N ratios of lake Karewa sediments at Romushi can also be interpreted along the above lines. The $\delta^{13}\text{C}$ values of the organic fraction of two grab samples from the modern lake Dal in Kashmir were found to be -24.9 and -25.7% , respectively. Although their C/N ratio was not determined, as the samples were collected from the lake margin at very shallow depths, the values support the contention that the relative depletion in ^{13}C could indeed be due to a greater contribution from terrestrial vegetation with typical average values of -27% , as for the C_3 -dominated biomass abundant in the valley today. The $\delta^{13}\text{C}$ values for typical lacustrine biomass using dissolved bicarbonate could be, on average, 10% heavier, depending on the hardness of the water, thus leading to values in the range of -17% (ref. 15). Variability in this range may result from differing relative contributions from terrestrial biomass and floating planktonic material, as well as from variation in other chemical parameters of the lacustrine environment¹⁸. As dramatic changes in these factors are required to induce large variations in the $\delta^{13}\text{C}$ record, a combination of $\delta^{13}\text{C}$ values with the C/N ratio of the sediments can lead to better resolution of the environmental record. C/N ratios as high as ~ 70 have been reported for terrestrial plants, those for aquatic plants average ~ 6 , and are rarely >10 (ref. 24). Therefore, assigning the extreme $\delta^{13}\text{C}$ value of -17% for the submerged aquatic plants, -27% for the floating plankton (assuming that they used only atmospheric CO_2) and terrestrial biomass, and C/N ratios of 6 and >10 for typical lacustrine and terrestrial biomass respectively^{24–26}, it is evident from Fig. 1 that the carbon source to the lake has varied in the past. Most significantly, there appear to have been at least three major episodes when contribution by terrestrial vegetation was dominant. These three phases can be identified as periods when the Karewa lake receded substantially or was at least reduced to very shallow conditions, probably caused by glaciations. As glacial periods in the tropics during the Quaternary have also tended to be dry, a model is proposed

whereby 'terrestrial' organic matter could contribute to the sediment by increased terrestrial plant growth on the exposed areas, or by emergent aquatics growing in parts of the lake exposed by a lowering of lake level. In at least two cases, the C/N ratios of samples representing this episode are not very much greater than 6, suggesting that, in addition to purely 'terrestrial' biomass, emergent aquatics may also have contributed to the total sediment. This hypothesis may be tested by organic geochemical studies and identification of chemical fossil markers²⁵. Note (Fig. 1) that these three phases occur at or very close to the magnetic boundaries marking the Reunion (N_4), Olduvai (N_5) and Jaramillo (R_7) transitions, suggesting that these phases occurred at ~ 2.01 , 1.87 and 0.73 Myr.

The low C/N ratios of most of the samples studied here indicate deep lake conditions (and hence warmer, wetter regimes), but their $\delta^{13}\text{C}$ values average $\sim -22\%$, which deviates from the extreme value of -17% expected of aquatics using dissolved bicarbonate. This 'lowering' of $\delta^{13}\text{C}$ values may be thought of as due to a decrease in the supply of molecular CO_2 caused by more flushing and a relative increase in floating planktonic material, or by less well understood processes such as methanogenesis¹⁸. The very low C/N ratios, much lower than 6, are also interesting. One explanation for low C/N ratios is an additional source of nitrogen to offset the aquatic C/N ratios, perhaps in the form of NH_4^+ ions, which tend to adsorb on to clay particles and are released during analysis²⁷. This, indeed, would be likely during warmer, wetter climates, when the lake received greater input from its drainage area, which was rich in dissolved NO_3 and NH_4^+ .

Assuming that the three peaks characterized by low $\delta^{13}\text{C}$ values and high C/N ratios indicate three major cold intervals, it is interesting to compare these with deep-sea records. Our relatively coarse sampling does not resolve all the intervals covered by deep-sea records, but a broad correlation appears to be present. Among the most complete marine sequences relevant for comparison are those provided in refs 28–30. The first of these comes from oxygen isotope studies of core

V-28-38²⁸, which extends nearly up to the Jaramillo subchron (0.97 Myr) and preserves the record of 22 glacial/interglacial cycles. The Brunhes/Matuyama boundary in this core marks the boundary of stages 19 and 20, an event probably correlatable with the lastest of the three low-lake-level regimes at Romushi. In the second detailed work, on ocean core V-28-179²⁹ and DSDP site 552A³⁰, covering the time interval up to 3.9 Myr, there is evidence of glaciation at the Matuyama/Olduvai boundary, followed by an oxygen isotope excursion suggesting warm conditions. This has a parallel at Romushi, where the warmer episode may be seen in 300 m of sediments that indicate deep-water conditions. The low Karewa lake level encountered at the Reunion magnetochron can also be correlated with the data from the above cores.

Received 18 February; accepted 17 June 1986.

- De Terra, H. & Paterson, T. T. *Studies on the Ice Age in India and Associated Human Cultures* (Carnegie Institute, Washington, 1939).
- Agrawal, D. P. et al. *Acta geol. hung.* **22**, 185-196 (1979).
- Singh, I. B. J. *palaeont. Soc. India* **27**, 71-110 (1982).
- Agrawal, D. P. in *Climate and Geology of Kashmir and Central Asia. The Last Four Million Years* (eds Agrawal, D. P., Kusumgar, S. & Krishnamurthy, R. V.) 1-12 (Today's & Tomorrow's, New Delhi, 1985).
- Burbank, D. W. & Johnson, G. D. *Nature* **298**, 432-436 (1982).
- Gupta, H. P. et al. in *Climate and Geology of Kashmir and Central Asia. The Last Four Million Years* (eds Agrawal, D. P., Kusumgar, S. & Krishnamurthy, R. V.) 101-108 (Today's and Tomorrow's, New Delhi, 1985).
- Krishnamurthy, R. V., DeNiro, M. J. & Pant, R. K. *Nature* **298**, 640-641 (1982).
- Dodia, R., Agrawal, D. P. & Vora, A. B. in *Climate and Geology of Kashmir and Central Asia. The Last Four Million Years* (eds Agrawal, D. P., Kusumgar, S. & Krishnamurthy, R. V.) 13-17 (Today's and Tomorrow's, New Delhi, 1985).
- Tandon, S. K., Pant, R. K. & Korisetar, R. *Man Envr.* **6**, 6-10 (1982).
- Burbank, D. W. & Johnson, G. D. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **43**, 205-235 (1983).
- Burbank, D. W. & Grant, M. J. Z. *Gletscherk Glazialgeol.* **21**, 229-236 (1985).
- Kusumgar, S., Agrawal, D. P. & Kotlia, B. S. in *Climate and Geology of Kashmir and Central Asia. The Last Four Million Years* (eds Agrawal, D. P., Kusumgar, S. & Krishnamurthy, R. V.) 13-17 (Today's and Tomorrow's, New Delhi, 1985).

These broad correlations with the deep-sea sediment record make the Karewa sequence a potentially powerful area for palaeoclimatic research. Closer sampling, better information on sedimentation rates and analysis from several nearby sections is necessary for a finer resolution of climatic changes in the valley and an understanding of their global significance.

We thank Professor D. P. Agrawal, Principal Investigator, MDPP, and Drs R. Korisetar, R. K. Pant, R. Ramesh and S. K. Tandon for help. The C/N ratios were measured at the Bhaba Atomic Research Centre, Bombay, in the laboratories of Professor M. C. Chaddha. This work was carried out as part of the Kashmir Palaeoclimate Project No. 12(23)79/SERC, funded by the Department of Science and Technology, Government of India.

- Kusumgar, S., Bhandari, N. & Agrawal, D. P. in *Climate and Geology of Kashmir and Central Asia. The Last Four Million Years* (eds Agrawal, D. P., Kusumgar, S. & Krishnamurthy, R. V.) 245-247 (Today's & Tomorrow's, New Delhi, 1985).
- Mankinen, E. A. & Dalrymple, G. B. *J. geophys. Res.* **84**, 615-626 (1979).
- Stuiver, M. *Quat. Res.* **5**, 251-262 (1975).
- Smith, B. N. & Epstein, S. *Pl. Physiol.* **47**, 380-384 (1971).
- O'Leary, M. H. *Phytochemistry* **20**, 553-567 (1981).
- Hakansson, S. *Quat. Sci. Rev.* **4**, 135-146 (1985).
- Buchardt, B. & Fritz, P. in *Handbook of Environmental Isotope Geochemistry: The Terrestrial Environment A* (eds Fritz, P. & Fontes, J.) 473-504 (Elsevier, Amsterdam, 1980).
- Nakai, N. *Proc. Japan Acad.* **48**, 516-521 (1972).
- Stiller, M. preprint.
- Nakai, N. preprint, Nagoya Univ. (1982).
- Nakai, N. et al. *Daiyinki-Kenkyu* (in the press).
- Meybeck, M. *Am. J. Sci.* **282**, 401-450 (1982).
- Leenheer, M. J. & Meyers, P. A. in *Advances in Organic Geochemistry, 1981* (eds Bjoroy, M. et al.) 309-316 (Wiley, London, 1983).
- Debyser, Y., Monvozier, L. J., Santvoit, D. & Daumas, R. in *Advances in Organic Geochemistry, 1981* (eds Bjoroy, M. et al.) 259-267 (Wiley, London, 1983).
- Milliman, D. J. & Quinchun, Xu *Am. J. Sci.* **284**, 824-834 (1984).
- Shackleton, N. J. & Opdyke, N. D. *Quat. Res.* **3**, 39-55 (1973).
- Shackleton, N. J. & Opdyke, N. D. *Nature* **270**, 216-219 (1977).
- Shackleton, N. J. et al. *Nature* **307**, 620-623 (1984).

Female pied flycatchers choose territory quality and not male characteristics

Rauno V. Alatalo, Arne Lundberg & Carolyn Glynn

Department of Zoology, Uppsala University, Box 561, S-751 22 Uppsala, Sweden

Female mate choice is central to the theory of sexual selection¹, but much of the evidence for the basis of this choice is inconclusive²⁻⁴. First, non-random mating may occur because of male-male competition³; and second, even if females can be shown to be choosing mates it is often difficult to tell exactly on what criteria they are choosing. Possible criteria include male genes⁵⁻¹⁰, parental ability³ and quality of the resources held by males^{11,12}, but in the wild these factors are often interdependent³. Here we report a field experiment with pied flycatchers that eliminates the correlation between male characteristics and territory quality. We find that age, size, plumage colour and song repertoire of males are all unimportant in female choice, and that territory quality is the single most important criterion.

The pied flycatcher (*Ficedula hypoleuca*, Muscicapinae), a hole-nesting, migratory, territorial bird species, can easily be attracted to nest boxes. It is sexually dimorphic in plumage, and males have a very complex song repertoire. Some males achieve polygyny by being polyterritorial^{13,14}. Males arrive on breeding territories about 1 week before females, male arrival order being correlated with female mating order. This mating pattern could be a consequence of the fact that many females arrive before all males have arrived so that only the first males are available as mates for the first females. However, among the available males, females still, on average, choose the males that

have arrived first (Fig. 1; shaded columns). Males vary remarkably in plumage from black to brown, and colour is partly age-dependent. Older males arrive first on the breeding grounds; these males are the blackest¹⁵, and seem to be preferred by females. Females may choose the first-arriving males (probably common in many bird species¹⁶) either because the 'best' male arrives first and mate choice is based on male quality (genes and/or parental ability); or because the first male to arrive takes the best territory and females choose males on basis of territory quality.

To separate these hypotheses it is necessary to distinguish between male characteristics (for example, arrival time, age, size and colour) and territory quality. We achieved this with a nest-box experiment which forces males to occupy territories at random.

We carried out field experiments from early May to mid-June 1985 at four deciduous woodlands near Uppsala, Sweden (59°50' N, 17°40' E). Each site consisted of two nearby (5-10 m) locations for boxes; we used new boxes with an entrance diameter of 28 mm to ensure high-quality nest sites and to reduce the variance in quality between sites. One site per woodland was drawn at random and the first nest boxes installed. Thus, the first male to arrive found a single location to occupy. As soon as he had taken that site and begun to sing, another pair of nest boxes was put up at the next randomly chosen site. No arriving male had a choice of territories because only a single vacant site existed at any one time. In this way we distributed 37 males at random sites relative to territory quality in the four woodlands.

Some females arrived and settled before the last males arrived. These females were all removed from their mates and all males resumed singing activity within 1 h. From the following day we noted the new pair order until all the males were mated. Then in two of the four study areas, we removed all females and allowed a second set to settle. For two woodlands, we used data from one set of females to measure female preference for a

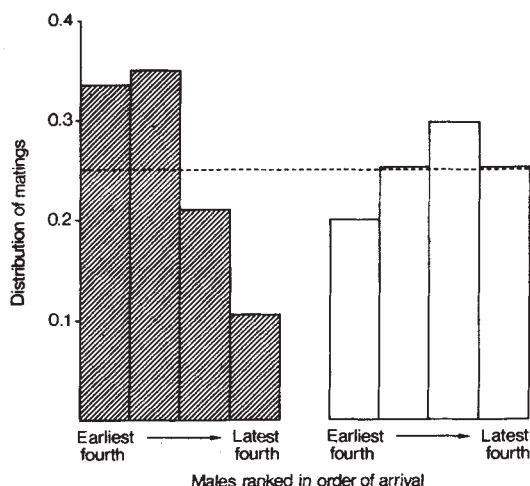


Fig. 1 Effect of male arrival order on mating success. Each day all unmated males were divided into four equally large groups according to their arrival, and matings were counted for each group and summed over all days. If females choose males irrespective of their arrival order, all groups should have equal proportions of matings (0.25). In unmanipulated areas there was a significant preference for early males ($\chi^2 = 35.56$, d.f. = 3, $P < 0.001$). In experimental areas, where the correlation between male arrival order and territory quality was eliminated by forcing the males to occupy nest sites in a random sequence, there was no significant difference from the random expectation ($\chi^2 = 0.91$, d.f. = 3, $P > 0.80$). The experiment and the control differ significantly ($\chi^2 = 10.87$, d.f. = 3, $P < 0.02$). Shaded areas, 1982-85 control, $n = 226$; white areas, 1985 experiments, $n = 37$; dotted line, random expectation. d.f. = degrees of freedom.

particular site; for the other two, we combined data from the two sets of paired females and used the average rank of female settlement order. In the latter two areas we also compared the settlement pattern of each set of females to determine whether females assessed males and/or territories in similar ways.

We captured, ringed and measured all birds. For males, we scored age and assessed colour by the percentage of brown or grey versus black feathers on the crown, mantle and back, measured wing and tarsus length and recorded weight. Samples of 20 songs from each male were recorded before they attracted females; pied flycatchers clearly reduce their singing after pairing¹⁷. From sonagrams we created a catalogue of notes to estimate male song repertoire size (a jack-knife estimate) and song phrase length. In statistical tests, we correlated female mating order with male characters and song performance (see Table 1). We calculated a standardized mating-order rank for males according to the formula

$$\frac{2 \times \text{rank} - 1}{2 \times n}$$

where rank is the mating order within areas and n is the total number of males in that area. The transformation gives a standardized rank ranging between 0 and 1, with an average of 0.5, allowing us to combine all four study areas into one data set.

Table 1 Spearman rank correlations between male characteristics and female mating order

Male characteristics	Range	r_s	n	P
Plumage darkness (%)	1-95	-0.047	37	0.81
Tarsus length (mm)	18.5-20.3	-0.069	35	0.69
Wing length (mm)	75.0-83.0	0.102	35	0.56
Weight (g)	11.8-13.8	-0.039	35	0.82
Age (yr)	1 or ≥ 2	0.056	32	0.76
Mean song phrase length (notes per song)	6.50-15.85	-0.081	36	0.64
Total repertoire size (unique notes per 20 songs)	6-71	-0.011	36	0.95

The two settling orders of females (after the removals) were significantly correlated (Spearman rank correlation $r_s = 0.782$, $P < 0.01$, $n = 15$) which shows that males were not chosen at random but in a similar order by both sets. Females clearly chose between males and visit several male territories before mating (R.V.A. and A.L., unpublished observations). There was no correlation between male arrival order and the order in which they obtained females ($r_s = -0.090$, $P = 0.60$, $n = 37$; see Fig. 1; white columns). Thus females did not choose early males *per se*, so we can directly compare female choice with male characteristics. The results show that no correlations exist between male characters; song length and repertoire size; and pairing order (Table 1).

Our conclusion is corroborated by earlier data on reproductive success of females mated with different males^{15,18}, which show that females do not benefit by choosing any particular male trait. Thus, we are left with territory quality alone as the main cue for female choice.

What determines the 'quality' of a territory? For the pied flycatcher, a good territory probably should include a nest site that is safe from predators¹⁹ and close to favoured feeding areas²⁰. In the experiment we eliminated much of the natural variations in territory quality, but we could still measure some variable characteristics of the nest sight and its surroundings. A multiple regression using these variables shows that a model including nest height (1.39-2.25 m), nest-tree circumference (0.40-1.63 m) and density of birches *Betula pubescens* (0-45% of area covered) correlates significantly with order of female choice ($F = 4.02$, $P = 0.02$, coefficient of determination $R^2 = 0.274$). Females prefer territories with a low density of birches and high nest sites in trees with thick trunks. Birches may be unfavourable for feeding whereas nest sites higher up in large trees are more safe against predators.

To evaluate the importance of nest-hole quality, we allowed half the males to pair with females in five study areas containing new and up to 2-year-old boxes, then we removed these females and replaced the original nest boxes with older ones (4-5 years old) with larger entrance holes (35 mm). The newer boxes of previously successful males were given to the unsuccessful group of males. When more females arrived, previously unsuccessful males became mated earlier (median rank 0.38, $n = 18$) than previously successful males (rank 0.63, $n = 17$, Mann-Whitney U -test $U = 73.5$, $P < 0.01$), indicating that nest-hole quality is important for female choice. Similarly, our studies of natural nest sites show that properties of the nest site, such as height and entrance size are important in determining territory occupation order and predation risk.

We thank A. Arak, N. B. Davies, D. B. A. Thompson, S. Ulfstrand and L. Wallin for constructive criticism. The work was supported by Swedish Natural Science Research Council grant NFR 2196-108.

Received 17 March; accepted 6 June 1986.

1. Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (Murray, London, 1871).
2. Halliday, T. R. in *Behavioural Ecology. An Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 180-213 (Blackwell, Oxford, 1978).
3. Halliday, T. R. in *Mate Choice* (ed. Bateson, P.) 3-32 (Cambridge University Press, 1983).
4. Partridge, L. in *Mate Choice* (ed. Bateson, P.) 227-255 (Cambridge University Press, 1983).
5. Zahavi, A. *J. theor. Biol.* **53**, 205-214 (1975).
6. O'Donald, P. *Genetic Models of Sexual Selection* (Cambridge University Press, 1980).
7. Lande, R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3721-3725 (1981).
8. Andersson, M. *J. Linn. Soc.* **17**, 375-393 (1982).
9. Hamilton, W. D. & Zuk, M. *Science* **218**, 384-387 (1982).
10. Kirkpatrick, M. *Evolution* **36**, 1-12 (1982).
11. Davies, N. B. in *Behavioural Ecology. An Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 317-350 (Blackwell, Oxford, 1978).
12. Askenmo, C. E. H. *Anim. Behav.* **32**, 972-980 (1984).
13. Alatalo, R. V., Carlson, A., Lundberg, A. & Ulfstrand, S. *Am. Nat.* **117**, 738-753 (1981).
14. Alatalo, R. V., Lundberg, A. & Ståhlbrandt, K. *Anim. Behav.* **30**, 585-593 (1982).
15. Alatalo, R. V., Lundberg, A. & Ståhlbrandt, K. *Behav. Ecol. Sociobiol.* **14**, 253-261 (1984).
16. Brooke, M. de L. *J. Anim. Ecol.* **48**, 21-32 (1979).
17. von Haartman, L. *Ibis* **98**, 460-475 (1956).
18. Alatalo, R. V., Gustafsson, L. & Lundberg, A. *Am. Nat.* **127**, 241-245 (1986).
19. Nilsson, S. G. *Ibis* **126**, 407-410 (1984).
20. Alatalo, R. V., Lundberg, A. & Ulfstrand, S. in *Habitat Selection in Birds* (ed. Cody, M. L.) 59-83 (Academic, Orlando, 1985).

Visual orientation in motion-blind flies is an operant behaviour

Reinhard Wolf & Martin Heisenberg

Institut für Genetik und Mikrobiologie, Röntgenring 11,
D-8700 Würzburg, FRG

Insect visual orientation is mostly considered as a stimulus-response phenomenon^{1,2}. This view may be challenged in *Drosophila melanogaster* by genetic dissection, as we show here. When suspended in a fixed position in a flight simulator³, which transforms torque into angular velocity of the panorama, *Drosophila* establishes a state of zero net rotation of the panorama (optomotor balance⁴). The genetic variant *rol sol*, with greatly reduced optic lobes, is entirely motion-blind but can respond with turning manoeuvres to the position of landmarks. Mutant flies maintain optomotor balance but do so irrespective of the sign of the visual feedback in the simulator, that is, whether angular velocity is inversely or directly proportional to torque. Using the amount, but not the direction, of pattern motion the flies stabilize the panorama by trying out which value of torque has the desired effect. We conclude that operant behaviour is a basic constituent of visual orientation in flies.

Visual orientation of dipteran flies has been investigated for 40 years, mainly with emphasis on stimulus-induced turning responses^{5,6}. Investigations rely on so-called open-loop experiments, in which the animal is partially immobilized so as to prevent its behaviour from interfering with the experimental stimuli. If, for instance, the head of a fixed flying fly is glued to the thorax, the stimuli reaching the eyes may be quantitatively assessed. In addition, the fly's yaw torque (the rotational momentum around the fly's vertical body axis) can be continuously recorded. Flies in this situation consistently modulate their yaw torque in response to pattern motion. In free flight such behaviour might serve to stabilize the retinal image and the direction of flight. It is this approach that has led to the idea of 'visual guidance', the notion that the orientation of a fly might be entirely understood as the result of its responses to the distribution and motion of patterns in its visual space.

We have studied the relation between visual input and yaw torque in *D. melanogaster* using genetic dissection of the visual system. Here, we report experiments with a genetic variant carrying, on the X chromosome, the two independent mutant genes *reduced optic lobes* KS221 (*rol*) and *small optic lobes* KS58 (*sol*)⁷. In *rol sol* flies the optic lobes have about 12% the original volume. The remaining parts are retinotopically organized with the normal number of ommatidia in the eye and columns in the neuropil.

No syndirectional modulation of yaw torque in response to a rotating striped drum is found in *rol sol* flies. This holds for various stripe widths, light intensities, pattern speeds and pattern contrast values. The flies seem to be entirely blind for the direction of motion (Fig. 1). Not so for the position of a landmark. Individual *rol sol* flies consistently and swiftly modulate yaw torque in response to the position of a rotating single stripe. Some flies try to turn towards the stripe, some away from it, some do not respond at all (Fig. 2). The individual responses are not invariant idiosyncrasies. If the same experiment with the same set of flies is repeated half a day later, non-responders may now respond and vice versa. Some flies have changed the sign of their response, others have not. The response seems to be under the influence of a slowly changing disposition which we cannot control. Statistically, *rol sol* flies more often try to turn away from the stripe than towards it.

In free flight yaw torque causes a fly to turn and thus induces visual motion of the retinal image in the opposite direction. Image motion, in turn, induces syndirectional yaw torque. This

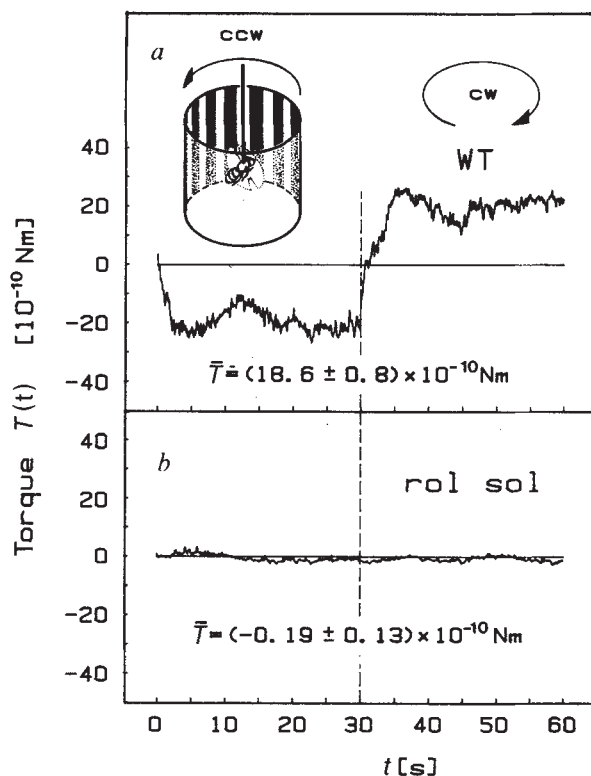
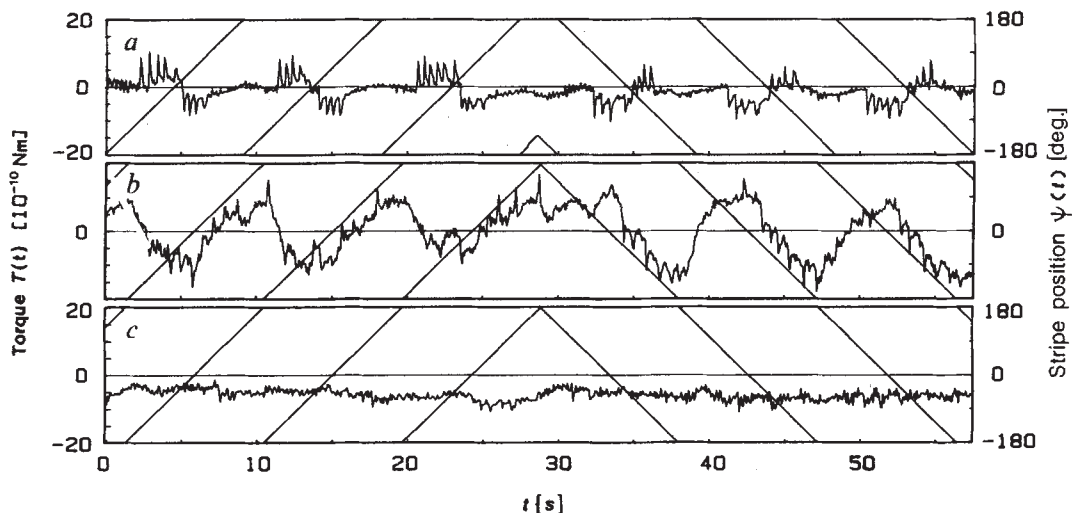


Fig. 1 Directional motion blindness of the mutant *rol sol*. Flies are immobilized by cooling and are glued with their heads and thorax to a bent silver wire (diameter 50 μ m) which for the experiment fits into the clamp of a torque meter. They are held horizontally in the centre of a vertical striped cylinder (pattern wavelength $\lambda = 18^\circ$) which rotates clockwise (cw) or counterclockwise (ccw) (contrast frequency $w/\lambda = 1$ Hz). The cylinder is homogeneously illuminated from behind (pattern luminance $\bar{I} = 10$ cd m⁻²). Males aged 2–4 days are used throughout. Torque traces are averages of 20 records of 10 wild-type (a) and 16 records of 8 *rol sol* (b) flies. Inset numbers ($\bar{T}$) represent mean and s.e.m. of the time-averaged response.

'feedback loop' situation can be imitated in stationary flight if the angular velocity of the panorama is made inversely proportional to the fly's yaw torque³. As soon as the loop is closed and the fly is allowed to control the panorama, wild-type *Drosophila* shifts its mean yaw torque to a level which causes zero net rotation of the panorama. This state is called optomotor balance and corresponds to a straight trajectory in free flight. One can test the fly's ability to maintain optomotor balance by superimposing on the fluctuations of the panorama an additional constant angular motion. *Drosophila* compensates this rotatory bias by an equivalent change in yaw torque⁴.

Using this test we find that to a small, but significant, degree *rol sol* flies are able to adjust their optomotor balance (Fig. 3a). How can a motion-blind fly, by its yaw torque modulations, control the angular velocity of the panorama? At first sight one would like to assume that the position-sensitive open-loop responses might account for the closed-loop behaviour. A fly which under open-loop conditions continuously generates turning manoeuvres away from the position of the stripe should, in the closed-loop situation, shift the stripe to a backward position and should keep it there. Thus each time the stripe would move to a lateral position a turning manoeuvre away from it would shift it back again. But this is clearly not how the flies do it. They can adjust their optomotor balance with the stripe at any position in the visual field. They generate turning manoeuvres towards and away from the stripe, even if a moment earlier, while the feedback loop was open, they responded in a stereotype fashion. Thus, the open-loop responses do not help us to understand the closed-loop-behaviour. We know from

Fig. 2 Yaw torque responses of individual *rol sol* flies with respect to a landmark. The cylinder carries a single grey stripe (width $\delta = 5^\circ$; height $H = \pm 40^\circ$) rotating around the fly with constant angular velocity $w = 40^\circ \text{ s}^{-1}$ (smooth trace, scale on the right). *a*, Some flies respond with intended turning manoeuvres away from the position of the stripe and keep this relation throughout the recording period. *b*, Other flies try to turn towards the stripe. *c*, Others show no significant response. Some time later the same flies may show different response characteristics. Note that no directionally motion-dependent responses are observed in these traces.



wild-type studies⁴ that flies respond differently under open- and closed-loop conditions.

Next, one might argue that in *rol sol* flies the behaviour is still mediated by directionally sensitive motion detectors which, however, are not operational under open-loop conditions. This explanation can be ruled out as well. As mentioned above, the directional responses of wild-type flies serve to stabilize the panorama. Inversion of the sign of the feedback (making the panorama rotate in the same direction as the intended turn of the fly) invariably destabilizes optomotor balance and eventually drives the stripe into permanent rotation. With *rol sol* this is not the case. Under both feedback conditions the mutant flies establish optomotor balance equally well (Fig. 3b). Even frequent switching between negative and positive feedback does not confuse them (Fig. 4).

We have investigated whether *rol sol* flies in the case of inverted feedback might learn in a fraction of a second to give up their 'phylogenetic knowledge' about the relation between torque and the direction of displacement in favour of the opposite relation. The time resolution of the torque traces in Fig. 4 is good enough to make this hypothesis highly unlikely. Learning would imply that after the inversion of the feedback the fly would systematically start out with the false manoeuvre. This is not observed. If present, the false response must be small compared with the continuous fluctuations in the torque trace which are not more than a few per cent of the fly's maximal torque modulations.

A more likely alternative to ultra-fast learning is that *rol sol* flies just do not display any directional responses which would destabilize optomotor balance under inverted feedback conditions. However, how can the flies find the torque level corresponding to zero net rotation of the panorama? The only way is by 'trying out', which means that the fly must initiate turning manoeuvres of both polarities and somehow stick for a longer time to those manoeuvres which give a better stabilization of the pattern.

Whatever the detailed mechanism of this trying out may be, the fly compares stripe positions at different times but evaluates only the amount, not the direction, of the displacement. This finding has an interesting implication regarding the relation between open- and closed-loop behaviour of *rol sol*. As Fig. 2 shows, the flies can use a landmark as a reference for their turning manoeuvres but apparently cannot, with significantly higher than 50% probability, generate the correct turning manoeuvre for shifting the stripe to a desired position. If they could, inversion of the sign of the feedback would drive the stripe into permanent rotation or to a fixed position. We suggest that *rol sol* flies, being motion-blind, have lost their 'phylogenetic

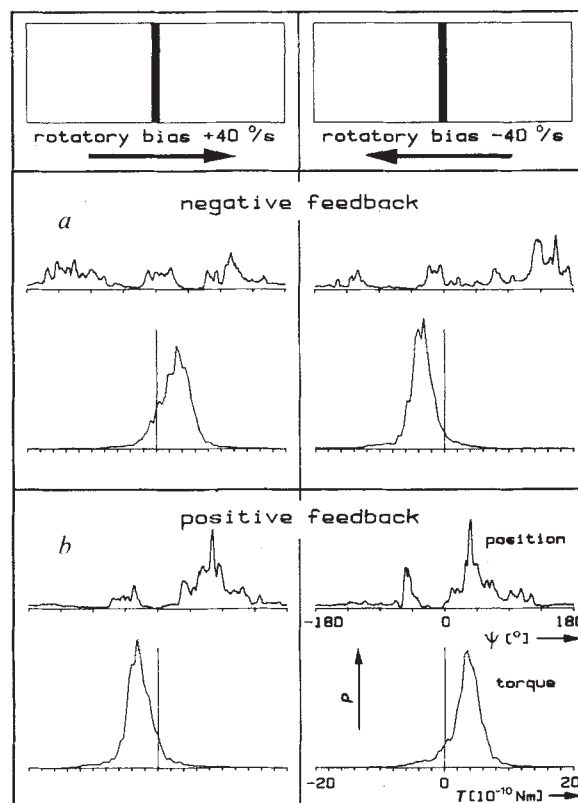


Fig. 3 Simulation of reafferent motion. Angular velocity of the panorama (a single vertical black stripe, $\delta = 10^\circ$, $H = \pm 40^\circ$) is proportional to yaw torque (closed-loop). *a*, Intended clockwise turning of the fly causes counterclockwise rotation of the stripe and vice versa (negative feedback). This is an imitation of free flight in which a clockwise turn causes a counterclockwise rotation of the retinal image of the surround. *b*, Here torque and pattern motion are in the same direction (positive feedback); a situation which never happens in free flight. Wild-type cannot stabilize the panorama while being in a positive feedback loop with the panorama. For *rol sol* flies the sign of the feedback does not matter. In both situations they compensate a rotatory bias corresponding to an angular velocity of the stripe of $w = +$ or -40° s^{-1} . The data are from 4×2 min of flight of a single *rol sol* fly. The upper diagrams are stripe position histograms and the lower diagrams are yaw torque histograms of the same records.

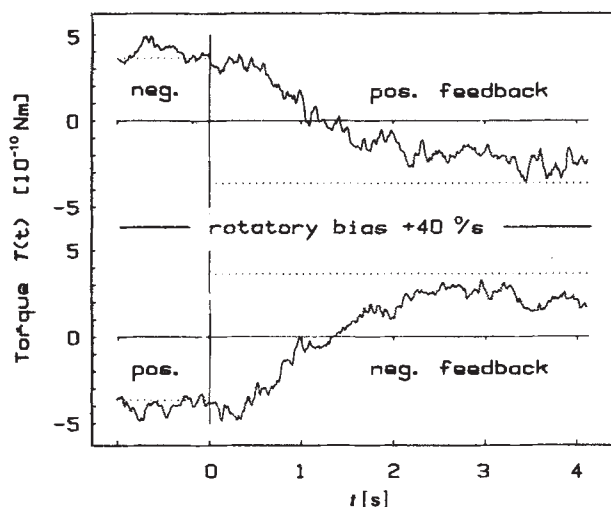


Fig. 4 No directional responses are involved in the stabilization of the panorama. Flies are in closed loop with a single stripe and have to compensate a rotatory bias of $w = +40^\circ \text{ s}^{-1}$ as in the experiment of Fig. 3. At arbitrarily chosen times (vertical lines) the sign of the feedback is inverted for 4.1 s. The torque traces show 1 s before and 4.1 s after the switch. Test trials are triggered when the stripe has a low angular velocity and a certain angular position. In the upper diagram the switch is from normal to inverted feedback, in the lower from inverted to normal. The switch causes a rotatory bias of $w = +80^\circ \text{ s}^{-1}$ with respect to the previous level of optomotor balance. Because of the prevalent tendency of the flies to generate turning manoeuvres away from the stripe in response to the onset of motion, care is taken to always include in the average traces pairs of test trials for the corresponding starting positions on the left and right side of the fly. Traces are averages of 132 (upper) and 152 (lower trace) test trials from 10 *rol sol* flies.

knowledge' of the (reafferent) visual effect of their turning manoeuvres.

The ability to cope with inverted visual feedback is not an exclusive characteristic of *rol sol* flies. We have shown earlier⁴ that in a special experimental situation in which the directionally motion-sensitive course control response is prevented from destabilizing orientation, wild type *Drosophila* is also able to adapt to these totally artificial conditions. Furthermore, in an electrophysiological study of muscles controlling wing beat amplitude in *Drosophila*, Götz⁸ discovered that one pair of muscles displays a similar flexibility.

The finding that motion-blind flies use an operant strategy for visual orientation suggests that in many circumstances orientation in insects, like in bacteria⁹ and humans¹⁰, is basically operant. No doubt, insects have evolved fast and reliable responses. However, goals, not responses, guide a fly; responses can be better understood as efficient tools for certain well-defined tasks.

We thank B. Bausenwein, E. Buchner, K. G. Götz and G. Pflugfelder for reading the manuscript, M. Esteve for correcting the English and A. Apffel for typing. This work was supported by Deutsche Forschungsgemeinschaft (He 986) and Universitätsbund Würzburg.

Received 21 March; accepted 16 June 1986.

1. Frankel, G. S. & Gunn, D. L. *The Orientation of Animals* (Dover, New York, 1961).
2. Wehner, R. in *Handbook of Sensory Physiology* Vol. VII/6c, 287-616 (Springer, Heidelberg, 1981).
3. Reichardt, W. & Wenking, H. *Naturwissenschaften* 56, 424-425 (1969).
4. Heisenberg, M. & Wolf, R. *Vision in Drosophila* (Springer, Heidelberg, 1984).
5. Götz, K. G. *J. comp. Physiol.* 99, 468-475 (1975).
6. Reichardt, W. & Poggio, T. Q. *Rev. Biophys.* 9, 311-375 (1976).
7. Fischbach, K. F. & Heisenberg, M. *Proc. natn. Acad. Sci. U.S.A.* 78, 1105-1109 (1981).
8. Götz, K. G. *Verh. dt. zool. Ges.* 83-99 (1983).
9. Adler, J. *Scient. Am.* 234, Pt. 4, 40-47 (1976).
10. Metzger, W. *Gesetze des Sehens* 2nd edn (Kramer, Frankfurt, 1953).

Elimination of action potentials blocks the structural development of retinogeniculate synapses

Ronald E. Kalil*, Mark W. Dubin†, Grayson Scott‡ & Louisa A. Stark†

* Department of Ophthalmology and Neurosciences Training Program, University of Wisconsin, 1300 University Avenue, Madison, Wisconsin 53706, USA

† Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, Colorado 80309-0347, USA

‡ Department of Anatomy, University of Wisconsin, Madison, Wisconsin 53706, USA

Although the influence of electrical activity on neural development has been studied extensively^{1,2}, experiments have only recently focused on the role of activity in the development of the mammalian central nervous system (CNS)³⁻¹⁰. Using tetrodotoxin (TTX) to abolish sodium-mediated action potentials^{11,12}, studies on the visual system show that impulse activity is essential both for the normal development of neuronal size⁵ and responsivity in the lateral geniculate nucleus (LGN)^{4,10}, and for the eye-specific segregation of geniculocortical axons^{3,7}. There have been no anatomical studies to investigate the influence of action potentials on CNS synaptic development. We report here the first direct evidence that elimination of action potentials in the mammalian CNS blocks the growth of developing axon terminals and the formation of normal adult synaptic patterns. Our results show that when TTX is used to eliminate retinal ganglion-cell action potentials in the cat from birth to 8 weeks, the connections made by ganglion cell axons with LGN neurones, retinogeniculate synapses, remain almost identical morphologically to those in the newborn kitten.

Retinogeniculate terminals in newborn kittens can be recognized easily by their round synaptic vesicles and unique, pale mitochondria^{13,14}. These terminals make simple, synaptic contacts almost exclusively with dendrites, via unusually large synaptic contact regions (Fig. 1a). By 8 weeks postnatal, retinogeniculate terminals are mature^{14,15}: on average, mean cross-sectional area is more than double that at birth, and the number of synaptic contacts is increased by more than 25% (Fig. 1b). Surprisingly, the average length of synaptic contacts decreases during development by ~30%. Thus, in the normal kitten there is a marked decline during development in the proportion of retinogeniculate terminal membrane specialized for synaptic apposition (Table 1).

The pattern of retinogeniculate synapses changes significantly during development in normal kittens as the result of two events, both of which begin at about 3 weeks and are largely completed by 8 weeks^{14,15}. First is the formation of synapses by retinogeniculate terminals with vesicle-containing profiles in the LGN. The origin of these F profiles, named for their flattened synaptic vesicles¹³, is not fully established, but many are probably specialized dendrites of inhibitory interneurons in the LGN¹⁶⁻¹⁸. Because the number of synaptic contacts made by retinogeniculate terminals with conventional dendrites is almost identical at birth and 8 weeks (Table 2), it is the addition of new synaptic contacts with F profiles that accounts for the overall increase in the number of contacts made by each retinogeniculate terminal during development. Second is the establishment of complex synaptic zones that consist typically of a centrally placed retinogeniculate terminal surrounded by postsynaptic dendrites and F profiles. Each complex zone is 'encapsulated' by glial processes which can isolate the retinogeniculate terminal and its postsynaptic constituents from the surrounding neuropil (Fig. 1b). By 8 weeks of age, ~30% of all retinogeniculate terminals are located within encapsulated zones (Table 2), and almost all of them make synaptic contacts

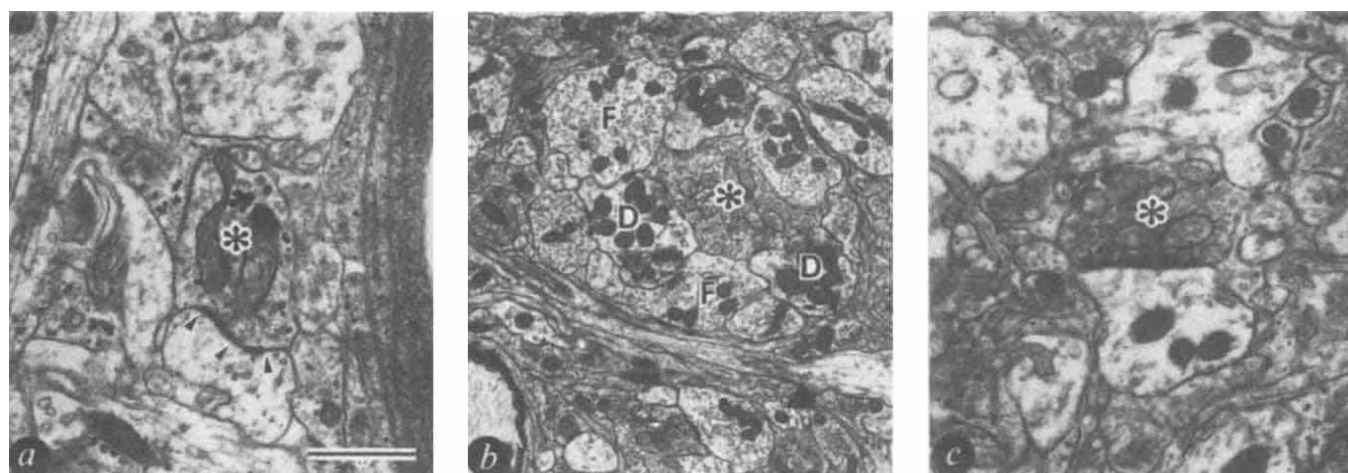


Fig. 1 *A*, Electron micrograph of a retinogeniculate terminal (asterisk) making an axodendritic synaptic contact, from a normal, newborn kitten. Note length of synaptic apposition (arrows). *B*, Retinogeniculate terminal (asterisk) surrounded by postsynaptic dendrites (*D*) and *F* profiles (*F*) in an encapsulated zone (see text) of a normal 8-week-old kitten. *C*, Retinogeniculate terminal (asterisk) from an 8-week-old kitten that had received intraocular injections of TTX from birth. This terminal makes an axodendritic contact and is very similar to the terminal shown in *A*. Bar, 1.0 μ m in *a* and *c*; 2.0 μ m in *b*.

Methods. Three groups of cats were studied: normal animals at one day ($n=2$) or 8 weeks ($n=3$) postnatal, cats given monocular injections of TTX on alternate days from birth to 8 weeks ($n=3$) to block all action potentials¹⁰, and controls given monocular injections of citrate buffer, the vehicle used for administering TTX, on alternate days from birth to 8 weeks of age ($n=2$). Impulse blockade was ensured in TTX-injected cats following an injection protocol developed from a separate series of physiological experiments (for details see Ref. 10). The lateral geniculate nuclei from all animals were prepared identically for microscopy according to standard procedures. Retinal whole mounts from each cat (excluding newborn) were stained for ganglion cell bodies with cresyl violet. Detailed measurements of retinogeniculate synaptic patterns, terminal size (cross-sectional area and perimeter) and synaptic contact length and number, and synaptic vesicle diameter were made from electron micrographs at magnifications of 12,000 $\times$, 25,000 $\times$ and 125,000 $\times$, respectively. The patterns of synaptic contacts made by retinogeniculate terminals were quantified on large-scale photographic montages by recording the identity of each profile postsynaptic to a terminal from the retina. The percentage of retinogeniculate terminals occurring in encapsulated zones was also noted. All measurements were made in lamina A of the LGN, which receives input only from the contralateral eye. In control cats injected with citrate buffer alone, retinogeniculate terminals only from the injected eye were analysed. In cats injected with TTX, retinogeniculate terminals from both eyes were measured, so that each animal served as its own control, permitting direct comparison between treated and untreated retinogeniculate terminals. The data from newborn kittens and normal 8-week-old kittens provided a reference for retinogeniculate synaptic development at times equivalent to the beginning and end of the TTX injections in experimental animals. Note that all measurements were made from micrographs produced from single thin sections, so they do not provide true estimates of the total number of synapses made by retinogeniculate terminals. Although this information can be obtained by complete serial reconstruction of retinogeniculate terminals¹⁸, this is unnecessary for the purpose of the present report. All relevant comparisons are based on relative differences, and these can be determined accurately by the sampling techniques used.

with both *F* profiles and dendrites^{14,15,19}. The remaining retinogeniculate terminals make synapses with dendrites or *F* profiles outside encapsulated zones.

In 8-week-old cats reared with TTX injections from birth, retinogeniculate terminals from injected eyes appear nearly identical to those in normal newborn kittens (Fig. 1c). The terminals are small, make mainly simple axodendritic contacts and their synaptic junctions are much longer than in normal 8-week-old cats. By contrast, retinogeniculate terminals from uninjected eyes look identical to those in normal 8-week-old animals, as do retinogeniculate terminals in cats that received control injections of citrate buffer.

These observations are confirmed in tables 1 and 2. Retinogeniculate terminals from TTX-injected eyes are very similar morphometrically to terminals from newborn kittens. Retinogeniculate terminals from uninjected eyes are statistically different from terminals connected to injected eyes, but are

equivalent to those in normal 8-week-old kittens, as are retinogeniculate terminals from control-injected kittens (Table 1). Measurements of synaptic vesicle diameter (data not shown) reveal no effect of TTX injections on vesicle size. Retinogeniculate terminals from TTX-injected eyes engage in synaptic patterns that resemble those of newborn kittens (Table 2). In particular, retinogeniculate terminals in both groups are almost never encapsulated, and rarely make synaptic contacts with *F* profiles. Terminals from uninjected eyes and from control-injected eyes develop normal synaptic patterns.

The retinas from TTX-injected kittens show no obvious pathology. However, the mean soma diameter of large ganglion cells (α cells) from injected eyes is reduced by 20% relative to corresponding cells from uninjected eyes. Although control experiments show that about half of this size reduction is caused by the injection procedure itself or by the citrate buffer, it seems that TTX retards the growth of retinal ganglion cells. This could

Table 1 Morphometric analysis of retinogeniculate (RLP) terminals and synapses

Animal	RLP terminals (T)		Length (L) (μ m)	RLP synaptic contact (S)		
	Perimeter (P) (μ m)	Area (μ m ²)		S/T	L/P	Sum of L/P
Control injected ($n=2$)	8.6 $\pm$ 0.2	3.6 $\pm$ 0.0	0.37 $\pm$ 0.02	1.87 $\pm$ 0.01	0.04 $\pm$ 0.01	0.08 $\pm$ 0.01
Uninjected ($n=3$)	8.9 $\pm$ 0.5	3.6 $\pm$ 0.1	0.39 $\pm$ 0.01	2.05 $\pm$ 0.20	0.04 $\pm$ 0.01	0.09 $\pm$ 0.01
TTX injected ($n=3$)	5.2 $\pm$ 0.2	1.5 $\pm$ 0.3	0.62 $\pm$ 0.07	1.32 $\pm$ 0.04	0.12 $\pm$ 0.01	0.16 $\pm$ 0.02
<i>P</i>	<0.01	<0.01	<0.05	<0.02	<0.01	<0.01
Normal newborn ($n=2$)	5.4 $\pm$ 0.2	1.4 $\pm$ 0.2	0.56 $\pm$ 0.08	1.41 $\pm$ 0.01	0.11 $\pm$ 0.0	0.16 $\pm$ 0.01
Normal eight weeks ($n=3$)	9.0 $\pm$ 0.2	3.8 $\pm$ 0.1	0.39 $\pm$ 0.03	1.81 $\pm$ 0.08	0.04 $\pm$ 0.01	0.07 $\pm$ 0.01

All measurements are from lamina A of the LGN. Means $\pm$ s.e.m., computed using the results from each cat as a single value. S/T, number of synaptic contacts per terminal; L/P, mean contact length as a percentage of terminal perimeter; Σ L/P, percentage of terminal perimeter occupied by synaptic contacts. All *p* values are derived from two-tailed *t*-tests, 4 d.f., and refer to the differences between the uninjected and TTX injected means in each column. Cats reared from birth to 8 weeks with unilateral or intraocular injections of TTX, or citrate buffer alone. For comparison, results from 2 newborn cats and 3 normally reared 8-week-old cats are shown.

Table 2 Retinogeniculate (RLP) synaptic patterns.

Animal	No. studied	RLP terminals		RLP synaptic contacts		
		Encapsulated (%)	With a synapse (%)	Average no. per RLP with Dendrites	F profiles	F:D
Control injected (n = 2)	94.0 ± 12.0	29.6 ± 1.5	70.0 ± 4.0	1.30 ± 0.04	0.44 ± 0.01	0.34 ± 0.01
Uninjected (n = 3)	97.3 ± 17.4	26.3 ± 2.6	75.8 ± 0.4	1.29 ± 0.03	0.46 ± 0.10	0.36 ± 0.08
TTX injected (n = 3)	89.7 ± 3.9	1.2 ± 1.2	70.9 ± 1.9	1.20 ± 0.01	0.08 ± 0.04	0.06 ± 0.03
P		< 0.001	NS	< 0.05	< 0.05	< 0.05
Normal newborn (n = 2)	68.5 ± 8.5	1.5 ± 0.2	87.0 ± 0.0	1.28 ± 0.08	0.08 ± 0.02	0.06 ± 0.02
Normal eight weeks (n = 3)	90.3 ± 21.9	30.5 ± 1.4	71.8 ± 0.9	1.30 ± 0.08	0.48 ± 0.05	0.37 ± 0.06

F:D, mean number of synaptic contacts made by RLP terminals with F profiles divided by the mean number of contacts made with dendrites. Other details as described in Table 1. NS, not significant.

reflect a reduction in the metabolic activity of silenced ganglion cells caused by the direct action of TTX on the retina. Alternatively, TTX could act at the axon terminals of ganglion cells and the reduction in cell body growth could be the result of an indirect, retrograde effect.

It is possible that axon terminals of silenced ganglion cells fail to develop because impulse blockade disturbs axonal transport, thereby depriving axon terminals of substances necessary for growth and development²⁰⁻²². This seems unlikely, however, as blocking axonal transport in the developing retinogeniculate system with colchicine does not produce morphological consequences similar to those described here²³. Another possibility is that TTX injections produce a loss of retinogeniculate terminals from the silenced eye which could lead to translaminal sprouting of normal retinogeniculate axons from adjacent layers of the LGN²⁴⁻²⁶. However, preliminary experiments using autoradiographic techniques to trace normal retinogeniculate axons in TTX-injected cats suggest that translaminal sprouting does not occur (R.E.K. and M.W.D., unpublished). Moreover, binocular injections of TTX, which should eliminate the possibility of translaminal sprouting, have the same effect on retinogeniculate synaptic development as monocular injections (R.E.K. and M.W.D., unpublished).

Our experiments demonstrate that when action potentials in the retinogeniculate pathway of the cat are blocked monocularly by intraocular injections of TTX, the morphological development of retinogeniculate axon endings is also blocked. This shows directly that electrical impulse activity is required during development for the normal structural maturation of retinogeniculate terminals and their contacts, as well as for synaptogenesis and the formation of complex synaptic patterns. The failure of these events to occur cannot be ascribed to generalized or systemic effects of TTX as retinogeniculate terminals connected to the uninjected eye developed normally.

Presumably, impulse activity influences the intracellular regulation of macromolecules necessary for neural development²⁷. Although it may prove difficult to attack problems at this level directly in the mammalian CNS, further insight concerning the mechanisms involved might be gained by substituting controlled, electrical stimulation for normal impulse activity⁷, or by manipulating the timing of TTX injections during development²⁸ to determine whether the collective effects on synaptic growth and patterning can be fractionated into separate developmental components.

We thank M. Behan, P. Spear, L. Stanford, L. Tong and A. Weber for helpful comments on the manuscript. Supported by NIH grants EY01331 (R.K.) and EY04629 (M.W.D.); NSF grants BNS76-00506, BNS79-14110 (M.W.D.); and grant 1-798 from the March of Dimes (M.W.D.).

Received 21 January; accepted 6 June 1986.

- Fawcett, J. W., O'Leary, D. D. M. & Cowan, W. M. *Proc. natl. Acad. Sci. U.S.A.* **81**, 5589-5593 (1984).
- Stryker, M. P. & Strickland, S. L. *Invest. Ophthalm. Vis. Sci. Suppl.* **25**, 278 (1984).
- Riccio, R. V. & Matthews, M. A. *Devl. Brain Res.* **19**, 173-182 (1985).
- Riccio, R. V. & Matthews, M. A. *Devl. Brain Res.* **20**, 55-68 (1985).
- Dubin, M. W., Stark, L. A. & Archer, S. M. *J. Neurosci.* **6**, 1021-1036 (1986).
- Narahashi, T., Moore, J. W. & Scott, W. R. *J. gen. Physiol.* **47**, 965-974 (1964).
- Catterall, W. A. *Rev. Pharmac. Toxicol.* **20**, 15-43 (1980).
- Guillery, R. W. *Z. Zellforsch. mikrosk. Anat.* **96**, 1-38 (1969).
- Kalil, R. E. & Scott, G. *Soc. Neurosci. Abstr.* **5**, 791 (1979).
- Mason, C. A. *Neuroscience* **7**, 541 (1982).
- Sterling, P. & Davis, T. L. *J. comp. Neurol.* **192**, 737-749 (1980).
- Fitzpatrick, D., Penny, G. R. & Schmechel, D. E. *J. Neurosci.* **4**, 1809-1829 (1984).
- Hamos, J. E., VanHorn, S. C., Raczkowski, D., Ulrich, D. J. & Sherman, S. M. *Nature* **317**, 618-621 (1985).
- Rapisardi, S. C. & Miles, T. P. *J. comp. Neurol.* **223**, 515-534 (1984).
- Grafstein, B. & Edwards, D. L. in *Axoplasmic Transport in Physiology and Pathology* (eds Weiss, D. G. & Gorio, A.) 21-26 (Springer, New York, 1982).
- Riccio, R. V. & Matthews, M. A. *Soc. Neurosci. Abstr.* **9**, 150 (1983).
- Edwards, D. L. & Grafstein, B. *Brain Res.* **299**, (1984).
- Matthews, M. A., Narayanan, C. H., Narayanan, Y. & Siegenthaler-Matthews, D. J. *Neuroscience* **7**, 405-422 (1982).
- Guillery, R. W. *J. comp. Neurol.* **146**, 407-420 (1972).
- Kalil, R. E. *Anat. Rec.* **172**, 339-340 (1972).
- Hickey, T. L. *J. comp. Neurol.* **161**, 359-382 (1975).
- Zigmond, R. E. & Bowers, C. W. *Rev. Physiol.* **43**, 673-687 (1981).
- Kalil, R. E., Dubin, M. W., Scott, G. L. & Stark, L. A. *Invest. Ophthalm. Vis. Sci. Suppl.* **26**, 287 (1985).

Noradrenaline hyperalgesia is mediated through interaction with sympathetic postganglionic neurone terminals rather than activation of primary afferent nociceptors

Jon D. Levine*, Yetunde O. Taiwo, Stephen D. Collins & Judy K. Tam

School of Medicine and Dentistry, University of California San Francisco, San Francisco, California 94143, USA

In hyperalgesic states, observed commonly as a major symptom of tissue inflammation or after central or peripheral nerve injury, non-noxious stimuli produce pain and noxious stimuli are perceived as more painful than usual. The mechanisms underlying the generation of hyperalgesia are not known. In patients with causalgia (burning pain and severe hyperalgesia after a nerve injury) activation of sympathetic post-ganglionic neurones or application of noradrenaline to painful skin exacerbates pain and hyperalgesia¹ while sympathectomy may afford complete relief². One suggestion is that noradrenaline released from sympathetic post-ganglionic neurones increases the discharge of damaged small-diameter afferents^{3,4} by a direct action on the primary afferents^{5,6}. Here we present a new model for noradrenaline-sensitive hyperalgesia and demonstrate that the site of action of noradrenaline is not on the primary afferents but rather is presynaptic on the sympathetic post-ganglionic terminals.

* To whom correspondence should be addressed at: Division of Rheumatology and Clinical Immunology, U-426 (Box 0724), University of California, San Francisco, San Francisco, California 94143, USA.

- Harris, W. A. *Rev. Physiol.* **43**, 689-710 (1981).
- Fawcett, J. W. & O'Leary, D. D. M. *Trends Neurol. Sci.* **8**, 201-206 (1985).
- Stryker, M. P. *Soc. Neurosci. Abstr.* **7**, 842 (1981).
- Archer, S. M., Dubin, M. W. & Stark, L. A. *Science* **217**, 743-745 (1982).
- Kupperman, B. D. & Kasamatsu, T. *Nature* **306**, 465-468 (1983).

The Randall-Selitto paw-withdrawal test⁷ was used to measure pressure nociceptive thresholds in male Sprague-Dawley rats (250–350 g). Hyperalgesia was defined as the percentage decrease in the pressure nociceptive threshold produced by a test agent. Chloroform (50 μ l) was applied daily to the dorsum of the hindpaw for 90 s to produce a hyperalgesic lesion⁸. Drugs or vehicles (10 μ l) were injected intradermally; drugs were dissolved in saline except indomethacin, which was dissolved in 2% sodium bicarbonate and then titrated to pH 7.2 with monosodium phosphate and prostaglandin E₂ which was dissolved in a vehicle of 10% ethanol in saline. Phentolamine, yohimbine and prazosin were injected at the same intradermal site 30 min before noradrenaline (NA) was injected. Indomethacin was injected intraperitoneally (i.p) 30 min before any other test substance.

Destruction of sympathetic post-ganglionic neurones (SPGNs) was produced by injecting rats either with guanethidine (20 mg per kg body weight per day for 6 weeks⁹) or 6-hydroxydopamine (6-OHDA, 75 mg per kg per day for 3 days intraperitoneally¹⁰), or via excision of the subdiaphragmatic paravertebral ganglia after abdominal incision.

Hyperalgesia produced by chloroform was not statistically significant when measured 5 hours after the first treatment but was at 24 h, just prior to the second treatment. Twenty-four hours after the second chloroform treatment the nociceptive threshold decreased further (Table 1) and then remained constant, even with further chloroform treatments.

Table 1 Pressure nociceptive-threshold after daily chloroform treatments

Effect of chloroform treatment	Change in nociceptive thresholds†
At 5 h	-0.2 $\pm$ 1.9%
At 24 h	-18.6 $\pm$ 2.9% (n = 12)
At 48 h	-26.9 $\pm$ 2.4% (n = 18)
Effect of indomethacin	
Control	
Indomethacin	+3.7 $\pm$ 2.8% (n = 18)
Chloroform treated	
Indomethacin	-23.1 $\pm$ 5.7% (n = 12)
+ noradrenaline	+7.7 $\pm$ 5.7% (n = 12)*
	-0.1 $\pm$ 1.4% (n = 12)

Threshold was tested 5 h after the first daily treatment and just prior to each subsequent application of chloroform. With additional daily treatment, after 48 h, no further decrease in threshold (that is increase in hyperalgesia) was observed to occur. Indomethacin, a cyclooxygenase inhibitor, caused no significant change in nociceptive threshold 30 min after its injection in the no-treatment control group. However, indomethacin caused complete reversal of the hyperalgesia induced by chloroform and prevented NA-induced reduction in nociceptive threshold seen in chloroform-treated rats that had intact sympathetic nervous systems (see Table 2). All intradermal injections, in this and subsequent tables, were in a volume of 10 μ l.

* $P < 0.001$ by a two-tailed Student's *t*-test.

† Mean $\pm$ s.e.m.

In control rats before chloroform treatment, neither intradermal injections of NA nor phentolamine (an α -adrenergic receptor antagonist) produced a significant change in nociceptive threshold (Table 2). In rats made hyperalgesic with chloroform, however, NA produced a further decrease in nociceptive threshold—that is the chloroform-induced hyperalgesia was significantly enhanced by NA. When we administered a high dose of NA (100 μ g), tenfold greater than was required to elicit marked hyperalgesia in chloroform-treated rats, to no-treatment control rats, we failed to produce hyperalgesia in the no-treatment controls. Phentolamine produced a dose-dependent increase in nociceptive threshold in chloroform-treated rats, and returned the nociceptive threshold to normal baseline at high doses (>1 μ g) (Table 2). Yohimbine, the relatively specific α_2 -antagonist, also abolished the hyperalgesic state induced by

Table 2 Nociceptive threshold 20 min after intradermal injection of adrenergic agents in control, chloroform-treated and chloroform-treated sympathectomized rats

Adrenergic agents in chloroform treated rats	
Control	
Noradrenaline (10 μ g)	-1.6 $\pm$ 4.3% (n = 12)
Noradrenaline (100 μ g)	-0.1 $\pm$ 2.0% (n = 12)
Phentolamine (10 μ g)	+2.2 $\pm$ 2.6% (n = 6)
Chloroform treated	
Noradrenaline (10 μ g)	-15.9 $\pm$ 2.1% (n = 18)*
Vehicle	+1.8 $\pm$ 2.1% (n = 6)
Phentolamine (0.01 μ g)	+15.7 $\pm$ 6.2% (n = 6)
Phentolamine (0.1 μ g)	+21.8 $\pm$ 4.0% (n = 16)*
Phentolamine (1 μ g)	+30.2 $\pm$ 3.3% (n = 24)*
Phentolamine (10 μ g)	+25.0 $\pm$ 3.3% (n = 18)*
Yohimbine (5 μ g)	+28.0 $\pm$ 4.7% (n = 24)*
Prazosin (10 μ g)	+2.6 $\pm$ 1.8% (n = 36)
Adrenergic agents in sympathectomized rats	
After chronic guanethidine sympathectomy	
Chloroform treatment	-0.1 $\pm$ 2.3% (n = 24)
+ noradrenaline (10 μ g)	+1.9 $\pm$ 3.2% (n = 24)
+ phentolamine (10 μ g)	+1.8 $\pm$ 1.7% (n = 14)
After chronic 6-hydroxydopamine sympathectomy	
Chloroform treatment	-3.9 $\pm$ 5.2% (n = 6)
+ noradrenaline (10 μ g)	+4.5 $\pm$ 2.2% (n = 6)
+ phentolamine (10 μ g)	+0.3 $\pm$ 2.5% (n = 6)
After chronic surgical sympathectomy	
Chloroform treatment	-2.5 $\pm$ 3.4% (n = 14)
+ noradrenaline (10 μ g)	+1.6 $\pm$ 2.3% (n = 14)
+ phentolamine (10 μ g)	+1.3 $\pm$ 1.7% (n = 14)

Intradermal injection of NA or phentolamine failed to produce a significant change in threshold in controls, but in rats treated with chloroform NA produced a significant decrease in threshold (increase in hyperalgesia). Phentolamine caused a dose-dependent reversal of chloroform-induced hyperalgesia. Yohimbine, a relatively α_2 -specific antagonist, also reversed chloroform-induced hyperalgesia while prazosin, a relatively α_1 -specific antagonist, caused no significant effect on nociceptive threshold in chloroform-treated rats. Values are expressed as change from the baseline threshold measured just prior to injection, of each test substance. In sympathectomized rats there was no significant change in nociceptive threshold produced by adrenergic agents.

* $P < 0.001$ by a two-tailed Student's *t*-test.

chloroform. Prazosin, a relatively specific α_1 -antagonist, was without effect. These data demonstrate that in the absence of injury, the nociceptive threshold is insensitive to NA and therefore suggests that NA acts in injured rats on a process not present in normal animals, namely a catecholamine-modifiable hyperalgesia. We hypothesize that the effect of phentolamine results from a presynaptic blockade of noradrenaline release from SPGNs¹¹ or blockade of the receptor on which the released noradrenaline acts.

To evaluate the hypothesis that SPGNs contribute to chloroform-induced hyperalgesia, we studied the effects of chloroform in groups of rats sympathectomized by guanethidine, surgery or 6-OHDA. Sympathectomy by any of these routes completely prevented chloroform-induced hyperalgesia (Table 2), indicating that SPGNs are a necessary element. We next tested whether NA would produce hyperalgesia in these sympathectomized chloroform-treated animals. Surprisingly, NA was without effect in sympathectomized rats, even after 7 days of chloroform treatment. Since chloroform did not produce hyperalgesia in sympathectomized rats and since exogenous NA could not elicit hyperalgesia in these animals, it appears that the target of the NA in enhancing hyperalgesia may be presynaptic receptors on the SPGN terminal.

NA may elicit the release of prostaglandins from SPGNs^{12–14}. Since prostaglandins are also thought to act directly on small-

diameter afferents, we set out to establish whether prostaglandins are involved in the NA-modifiable chloroform-induced hyperalgesia. In fact, indomethacin reversed chloroform-induced hyperalgesia and completely blocked the exacerbation of hyperalgesia produced by intradermal injection of noradrenaline in chloroform-treated rats (Table 1). Since another model of hyperalgesia, that induced by bradykinin, is also mediated by prostaglandin release¹⁵, we next tested whether bradykinin-induced hyperalgesia is also dependent on intact SPGNs. Consistent with this hypothesis, we found that intradermal injection of bradykinin (100 ng) produced hyperalgesia in untreated rats but failed to produce hyperalgesia in rats sympathectomized with guanethidine (Table 3).

Beside increased prostaglandin release, an induction of increased sensitivity of nociceptor afferents to prostaglandins could produce a similar result. The latter possibility was studied with exogenously administered prostaglandins in control and chloroform treated rats. To eliminate endogenous prostaglandin production, indomethacin was given to both control and chloroform treated animals 30 minutes prior to prostaglandin injection. Under these conditions the hyperalgesic response to exogenous prostaglandins was similar in chloroform and control animals at all prostaglandin doses tested (Table 3). Thus, sensitivity of the primary afferent nociceptor to prostaglandins is not altered by chloroform treatment.

Table 3 Change in nociceptive threshold 20 min after intradermal injection of bradykinin or PGE2

Effect of bradykinin	
<i>Bradykinin</i>	
Control	-20.4 ± 3.1% (n = 12)*
Sympathectomized	+1.8 ± 1.9% (n = 12)
Exogenous prostaglandin in indomethacin-treated rats	
<i>Control</i>	
Prostaglandin E ₂ (0.1 ng)	-2.7 ± 4.0% (n = 6)
(1.0 ng)	-16.6 ± 4.0% (n = 6)*
(10 ng)	-28.1 ± 2.9% (n = 6)*
(100 ng)	-31.1 ± 2.5% (n = 6)*
<i>Chloroform</i>	
Prostaglandin E ₂ (0.1 ng)	-2.8 ± 5.2% (n = 6)
(1.0 ng)	-12.8 ± 6.8% (n = 6)*
(10 ng)	-25.2 ± 6.4% (n = 6)*
(100 ng)	-22.1 ± 8.1% (n = 6)*

Bradykinin elicited a significant decrease in nociceptive threshold in controls but not in sympathectomized rats. PGE2 produced a dose-dependent hyperalgesia in controls and also in chloroform-treated rats. The latter two groups of rats were pretreated with indomethacin, 30 min prior to injection of PGE2.

* $P < 0.001$ by two-tailed Student's *t*-test.

Our findings show that chloroform can induce an alteration in peripheral tissues such that ongoing activity in SPGNs now sensitizes primary afferent nociceptors, producing hyperalgesia. NA, bradykinin and possibly other inflammatory mediators, when applied or present in chloroform-treated skin elicit a prostaglandin-mediated hyperalgesia probably via action on presynaptic terminals of the SPGN. The source of the NA-evoked prostaglandin release is unknown. One possibility is that catecholamines, by interaction with receptors on SPGN terminals release prostaglandins from the terminals of the SPGNs, as has been proposed for other systems¹²⁻¹⁴. Prostaglandins would, in turn, mediate the hyperalgesic effect of the noradrenaline via direct action on small-diameter afferents.

The mechanism suggested by the present experiments can also account for some unexplained clinical observations. In the skin of patients with causalgia, there is a delay of several minutes in the onset of NA-evoked hyperalgesia¹⁶. Although this is unexpected if the NA were acting directly on cut afferents, it is

consistent with the time course of prostaglandin-induced hyperalgesia¹⁷. The present study also may explain why corticosteroids as well as sympatholytic agents are clinically effective in the treatment of causalgia, a syndrome due to hyperactivity of SPGNs, since corticosteroids suppress prostaglandin synthesis. Corticosteroids may act similarly in suppressing the activity of small-diameter afferents in an experimentally induced neuroma¹⁸.

This work was supported by NIH grants NS21642, AM32634 and DE05369. J.D.L. is a Hartford Foundation Fellow. We are grateful to Professors Allan Basbaum and Roger Nicoll and Drs Philip Heller and Chauncey Bowers for many helpful discussions.

Received 21 April; accepted 23 June 1986.

1. Wiesenfeld-Hallin, Z. & Hallin, R. G. *Hum. Neurobiol.* **3**, 41-46 (1984).
2. Loh, L. & Nathan, P. W. *J. Neurol. Neurosurg. Psychiat.* **41**, 664-671 (1978).
3. Wall, P. D. & Gutnick, M. *Expl. Neurol.* **43**, 580-593 (1974).
4. Devor, M. & Janig, W. *Neurosci. Lett.* **24**, 43-47 (1981).
5. Devor, M. *J. Auton. Nerv. Sys.* **7**, 371-384 (1983).
6. Blumberg, H. & Janig, W. *Pain* **20**, 335-354 (1984).
7. Randall, L. O. & Selitto, J. J. *Arch. Int. Pharmacodyn.* **3**, 409-419 (1957).
8. Fitzgerald, M. J. T., Folan, J. C. & O'Brien, T. M. *J. invest. Dermat.* **64**, 169-174 (1975).
9. Burnstock, G., Evans, B., Gannon, B. J., Heath, J. W. & James, V. *Br. J. Pharmac.* **43**, 295-301 (1971).
10. Finch, L., Haeusler, G. & Thoenen, H. *Br. J. Pharmac.* **47**, 249-260 (1973).
11. Starke, K., Wagner, J. & Schumann, H. *J. Arch. Int. Pharmacodyn.* **195**, 291-308 (1977).
12. Stjärne, L. *Acta physiol. scand.* **86**, 574-576 (1972).
13. Ramwell, P. W., Shaw, J. E. & Kucharski, J. *Science* **149**, 1390-1391 (1965).
14. Westfall, T. C. *Physiol. Rev.* **57**, 659-728 (1977).
15. Juan, H. & Lembeck, F. *Agents Actions* **6**, 642-645 (1976).
16. Wallin, B. G., Torebjörk, E. & Hallin, R. G. in *Sensory Functions of the Skin of Primates with Special Reference to Man* (ed. Zotterman, Y.) 489-499 (Pergamon, Oxford, 1976).
17. Ferreira, S. H., Nakamura, M. & Saletto de Abreu Castro, M. *Prostaglandins* **16**, 31-37 (1978).
18. Devor, M. *Pain* **22**, 127-137 (1985).

A physiological role for titin and nebulin in skeletal muscle

Robert Horowitz*, Ellis S. Kempner†, Margaret E. Bisher† & Richard J. Podolsky*

* National Institute of Arthritis and Musculoskeletal and Skin Diseases and † National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

Production of active force in skeletal muscle results from the interaction of myosin-containing thick filaments with actin-containing thin filaments. These muscles are also passively elastic, producing forces that resist stretch independently of ATP splitting or of interaction between the filaments. The mechanism of this passive elasticity is unknown; suggestions include gap filaments in the region between thick and thin filaments in muscles stretched beyond filament overlap¹⁻⁵, or intermediate filaments which connect successive Z-disks⁶. Recently, the two exceptionally large proteins titin (also called connectin) and nebulin (originally called band 3) have been implicated in passive elasticity (for review see refs 7, 8). Here, we show that after these proteins are degraded by low doses of ionizing radiation, the ability of single skinned muscle cells to generate both passive tension in response to stretch and active tension in response to calcium is greatly reduced. These effects are accompanied by axial misalignment of thick filaments. Titin and/or nebulin apparently provide axial continuity for the production of resting tension on stretch and also tend to keep the thick filaments centred within the sarcomere during force generation.

Titin migrates as two chemically and immunologically similar bands on SDS polyacrylamide gels. The more slowly migrating band is more prominent, and estimates of the relative molecular mass (M_r) of the heavier component range from 1,400,000 to

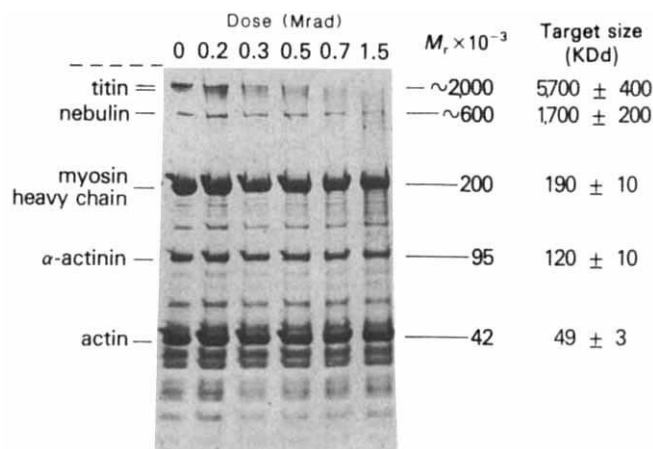


Fig. 1 SDS polyacrylamide gel electrophoresis of homogenates from control and irradiated skinned fibres. M_r s of myosin heavy chain, α -actinin and actin are from published values¹⁸. Estimates for the molecular weight of the heavier band of titin range from 1400 to 2800 K (refs. 8 and 9, respectively) and for nebulin from 500 to 800 K (refs. 10 and 9, respectively). Strips of rabbit psoas muscle were chemically skinned by the method of Wood *et al.*²⁵ in a solution containing 150 mM potassium propionate, 5 mM KH_2PO_4 , 3 mM magnesium acetate, 5 mM EGTA and 3 mM Na_2ATP , pH 7.0, at 5 °C. They were then immobilized near slack length and soaked for 10–15 min in skinning solution containing 15% sucrose. The skinned muscle strips were then rapidly frozen in isopentane cooled with liquid nitrogen and irradiated with 13 MeV electrons at -135°C ¹⁵. Following irradiation the fibres were thawed in skinning solution at 5 °C and homogenized in 20 mM Tris and 2 mM EDTA, pH 8.0, using a Dounce homogenizer. The homogenates were diluted 1:1 with 2% SDS, 20% glycerol, 80 mM dithiothreitol and 0.1% bromophenol blue, pH 8.0, boiled for 1–2 min and then electrophoresed by the method of Somerville and Wang²⁶. The slab gels were stained with Coomassie blue R-250 and staining was quantitated with a Beckman DU-8 scanning densitometer. Peak areas were determined for samples from fibres that had absorbed 0–18 Mrad. Target sizes are derived from the linear regression of $\ln$ (mean peak area) versus dose. In each case the slope and its standard error were determined and used to calculate target size and standard error of the target size, respectively.

2,800,000 (1,400 K–2,800 K)^{8,9}. Nebulin, of $M_r \sim 600$ K, is located in the I band where it exhibits an elastic stretch dependence, as it maintains the same proportional distance between the Z line and M line as the sarcomere is stretched¹⁰. Titin is localized in both the A band and I band, but only the I-band domain exhibits elastic stretch dependence^{11–13}.

Previous hypotheses of functions for titin and nebulin have relied primarily on structural evidence. Observation of contractile function in muscle fibres in which these proteins have been selectively destroyed or removed would reveal whether they have a physiological role in force generation. Taking advantage of the large size of these proteins, we have accomplished this by radiation inactivation, a technique which consists of exposing frozen biological material to ionizing radiation and observing the dose-dependency of the loss of biological function. This loss can be described by target analysis, which for a single target predicts an exponential decline with increasing radiation dose: $\ln(A/A_0) = -kD$, where A/A_0 is the fraction of surviving activity, D is the dose expressed in rads and k is a constant¹⁴. The analysis is valid because a single hit by radiation imparts enough energy to break many covalent bonds and to denature a peptide totally. As radiation hits occur randomly throughout the mass of the material, larger peptides have a greater probability of being destroyed than smaller ones. The rate of loss of activity with increasing dose, k , is therefore directly related to the mass of the functional unit, or target, by the following relation¹⁴: $M_r = (6.4 \times 10^{11})kS_r$. The coefficient of 6.4×10^{11} is derived from classical target analysis¹⁴ and S_r is an empirically determined correction to account for the decreased sensitivity of biological molecules to radiation at low temperature¹⁵ ($S_r = 2.8$ at -135°C). Because the radiation sensitivity of macromolecules is directly related to their mass, this technique can be used to study the function of the largest components involved in specific cellular processes.

We irradiated strips of skinned rabbit psoas muscle and observed the destruction of individual polypeptides using SDS

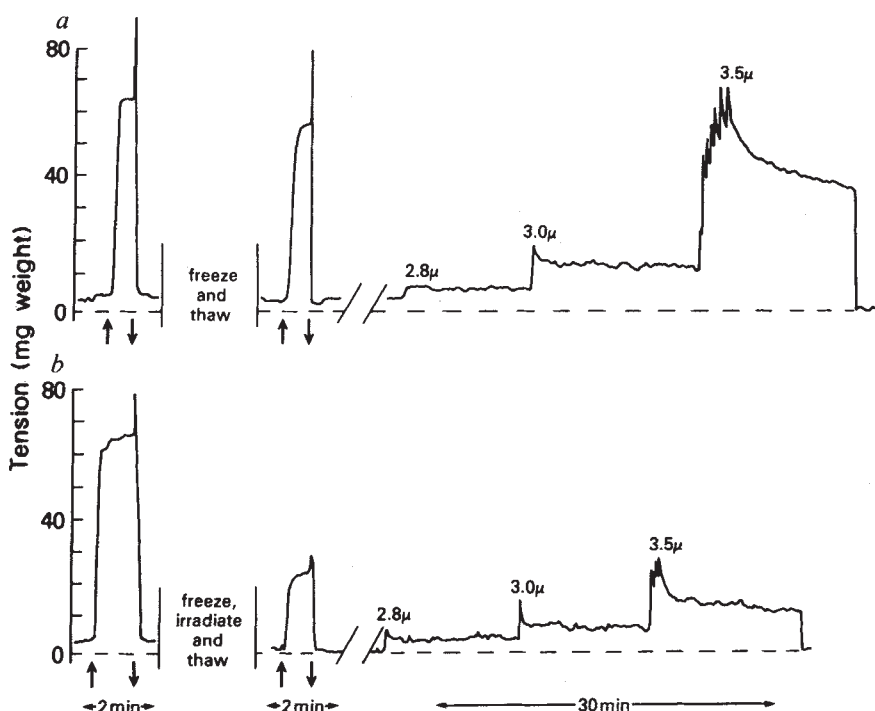


Fig. 2 Protocol for mechanical experiments. *a*, Dose = 0 Mrad; *b*, dose = 0.5 Mrad. Bundles of skinned fibres were prepared as described in Fig. 1 legend. Single fibres were dissected and connected to a leafspring force transducer²⁷ and a fixed end using aluminum foil T-shaped clips²⁸. The solution bathing the fibre could be rapidly changed and its temperature thermoelectrically controlled using a bath similar to that described previously²⁹. All experiments were conducted at 7 °C. Sarcomere length was monitored by laser-light diffraction and set at 2.6 μm for active force measurements. Active force was measured in the presence of calcium by removing the fibre from relaxing solution and immersing it in activating solution at the time indicated by $\uparrow$, the fibre being returned to relaxing solution at time $\downarrow$. The fibre was then removed from the force measuring apparatus and pinned to a piece of styrofoam near its slack length to be frozen, irradiated and thawed as detailed in Fig. 1. After remounting the fibre in the apparatus used for measuring isometric force, the maximum calcium-activated force was again measured at a sarcomere length of 2.6 μm . Passive tension was then measured by stretching the fibres to various sarcomere lengths as indicated above the tension records. These values were normalized relative to the maximum calcium-activated force which had been measured before freezing. Tension

baseline was determined by releasing the fibre at the end of each experiment. Freezing and thawing alone decreased active tension to $78 \pm 2\%$ of its control level, but did not significantly affect passive tension. Relaxing solution consisted of 100 mM potassium propionate, 3 mM EGTA, 25 mM imidazole, 7 mM MgCl_2 , 5 mM Na_2ATP , 15 mM Na_2 -creatine phosphate and 20 U ml^{-1} creatine kinase, pH 7.0. Activating solution consisted of 3.5 mM CaCl_2 added to the relaxing solution to yield a final pCa of 4.5.

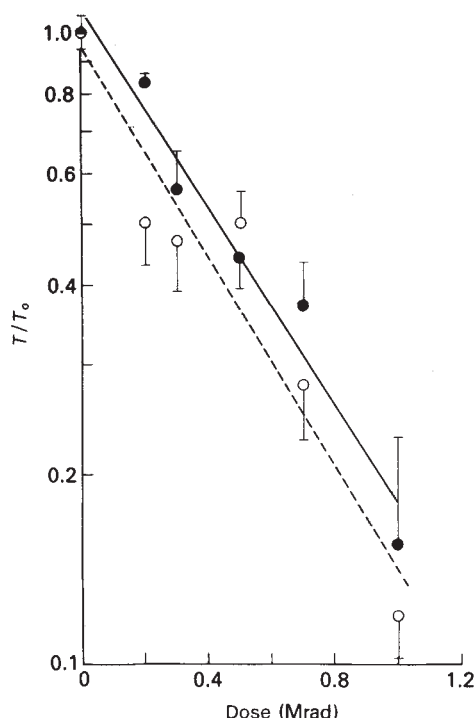


Fig. 3 Active and passive tension versus radiation dose. Filled circles, active tension, target size 3,200KDa; open circles, passive tension, target size 3,400KDa. Active and passive tensions were measured at sarcomere lengths of 2.6 μm and 3.0 μm , respectively. Passive tension was measured 10 min after stretch. Tensions (T) are expressed as a fraction of those measured in unirradiated controls (T_0). Bars indicate s.e.m. for 2–11 fibres.

polyacrylamide gel electrophoresis. As radiation dose is increased, the titin doublet and the nebulin band decrease in intensity (Fig. 1). Both proteins almost completely disappear after 1.5 Mrad of radiation, whereas at least 85% of myosin heavy chain, actin and α -actinin are intact after 1.5 Mrad of radiation.

Previously published estimates of the relative molecular mass of titin and nebulin were derived from gel electrophoresis of the denatured and dissociated peptides^{8–10}. Here we observe target sizes 2–4 $\times$ greater than these estimates (Fig. 1), suggesting that individual titin and nebulin peptides are linked together *in situ* (for example, by disulphide bonds) to allow transfer of radiation energy between individual peptides.

We studied force generation in single skinned rabbit psoas fibres using the protocol illustrated in Fig. 2. We measured active force by maximally stimulating the myofilaments with exogenous calcium and passive tension by stretching the resting fibre. At sarcomere lengths up to $\sim 3.5 \mu\text{m}$ most of the resting tension in frog semitendinosus muscle originates in the myofibrils^{16,17}. This is probably also true in rabbit psoas muscle, so that the magnitude of passive tension in the skinned fibre preparation is likely to reflect that of the intact muscle.

The slack sarcomere length of rabbit psoas fibres is near 2.2 μm and is not altered by exposure to radiation. In normal fibres stretched to a sarcomere length of 2.6 μm the resting tension was small but detectable (Fig. 2). However, in irradiated fibres resting tension was generally not measurable at 2.6 μm , indicating that radiation reduces the small amount of resting tension that is normally present at this length.

A dose of 0.5 Mrad significantly reduces both the calcium-activated force and the resting tensions measured 10 min after stretch to longer sarcomere lengths (Fig. 2). The dose-dependent decay of active and passive tension is illustrated in Fig. 3. Active and passive forces are both $<15\%$ of their control levels after 1 Mrad of radiation. Each of these processes decays exponentially over the same range of doses that is effective in destroying

titin and nebulin, radiation exposures which leave the major molecular components of thick filaments, thin filaments and Z-disks $>90\%$ intact.

The target size for the process of active force generation at a sarcomere length of 2.6 μm is $3,200 \pm 300\text{KDa}$ (kilodaltons) (calculated from the slope and standard error of the slope in Fig. 3). The target sizes for passive tension measured at sarcomere lengths of 2.8, 3.0 and 3.5 μm are $2,300 \pm 600\text{KDa}$, $3,400 \pm 500\text{KDa}$ and $4,400 \pm 500\text{KDa}$, respectively. Therefore, both physiological function and titin and nebulin (Fig. 1) all decay in a manner consistent with megadalton-sized targets. In contrast, myosin heavy chain, actin and α -actinin decay with target sizes ranging from 40 to 200KDa (fig. 1), in agreement with their published relative molecular masses¹⁸.

The target sizes for force generation lie between those for the destruction of titin and nebulin. Clearly there is not a simple one-to-one correspondence between the destruction of individual titin or nebulin molecules and the decrease in passive or active force. Such a correspondence is expected only if each titin or nebulin molecule is an independent, force-supporting unit arranged in parallel with all other titin and nebulin molecules. Obviously, many possibilities for more complex structural arrangements exist.

We examined the ultrastructure of normal and irradiated skinned fibres (Fig. 4). Irradiated relaxed fibres that had been fixed after stretch exhibited well-defined H zones, that is the central region of the A band which contains thick filaments but not thin filaments. The length of the H zones increased as the sarcomere length was increased from 2.6 to 3.0 μm , indicating that irradiation leaves the thin filaments intact and mechanically coupled to Z disks. However, after irradiation the edges of the A bands are no longer as sharply defined, and this is accompanied by a disordering of the M band at the centre of the sarcomere, indicating misalignment of thick filaments.

Muscle fibres fixed during calcium activation exhibit more disorder than resting muscle, but M lines and I bands are still clearly visible in the unirradiated samples. In contrast, calcium activation of irradiated fibres markedly amplifies the misalignment of thick filaments that was already present in the resting state. In some sarcomeres this leads to an almost complete disappearance of the I band as individual thick filaments are actively driven towards the nearest Z disk. Although still present, the M regions are also more disordered than in the unirradiated active muscle, but the A bands are more disordered than the M lines. Isolation of intact A bands has indicated that the M line may play a part in holding groups of thick filaments together¹⁹, but our results suggest that the M line is a structure through which individual thick filaments can slide, as previously observed in studies of highly stretched beef muscle³. These findings indicate that radiation destroys a structure responsible for centring thick filaments within the sarcomere during both passive and active force generation.

The A-band disorder observed in irradiated fibres is unlikely to be caused by breakage of thick filaments. If destruction of a single myosin molecule were sufficient to break the thick filament, then active force generation would exhibit a target size corresponding to the mass of an entire thick filament ($\sim 100,000\text{KDa}$). However, we observe a much smaller target size, indicating that more than one monomer must be destroyed to cause a break in the thick filament. Because gel studies reveal that individual myosin heavy chains are destroyed by single radiation hits, the breakage of a myosin filament must be described by a multiple target curve which exhibits a shoulder at low radiation doses²⁰. The radiation inactivation of active force shows no indication of a shoulder (Fig. 3). Therefore, thick-filament fragmentation is inconsistent with the mechanical data.

The biochemical, mechanical and ultrastructural evidence presented here suggests that a single structure containing titin and/or nebulin provides axial continuity for the production of

Fig. 4 Electron micrographs of control and irradiated fibres fixed at rest or during activation. *a*, Relaxed, sarcomere length 3.0 μm ; *b*, activated, sarcomere length 2.6 μm . Each panel shows a field of sarcomeres below which is a higher magnification view of a single sarcomere. Upper panels are unirradiated controls, lower panels have absorbed 0.5 Mrad. An A band, I band, H zone, M line and Z line are labelled in the upper panel of *a*. Bundles of skinned fibres ~ 0.25 mm in diameter were soaked in skinning solution with 15% sucrose added and then rapidly frozen in isopentane cooled with liquid nitrogen, then irradiated and subsequently thawed in skinning solution. The bundles were held isometrically during fixation and transferred to a bath containing fixative either 10 min after stretch to the indicated sarcomere length (*a*) or after being maximally activated with calcium ($p\text{Ca}=4.5$; *b*). The bundles were fixed for 2 h at 7°C in 2.5% glutaraldehyde added to either the relaxing or activating solution, then post-fixed for 1 h with 2% osmium tetroxide. The samples were dehydrated with ethanol and embedded in a mixture of Araldite and Epon. Gold or silver sections were cut with a diamond knife and stained with uranyl acetate and lead citrate. The stained sections were viewed and photographed using a Philips EM 300 microscope.

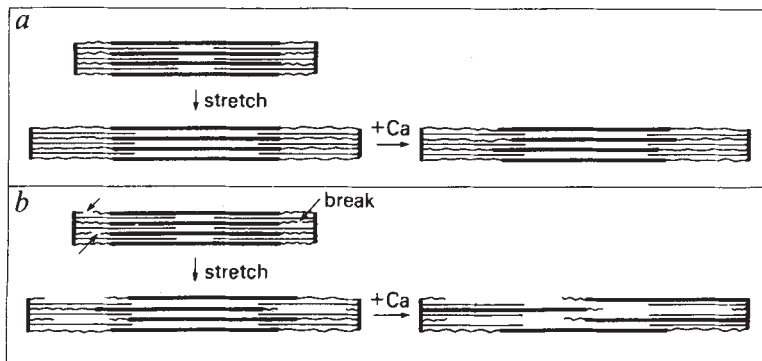
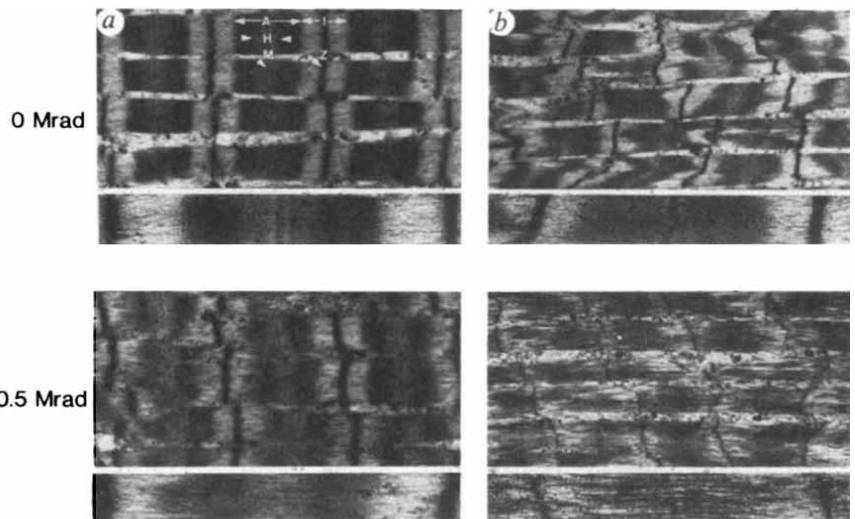


Fig. 5 Model sarcomere in which each end of the thick filament is linked to the nearest Z disk by elastic filaments made of titin and/or nebulin. *a*, Control; *b*, irradiated. Breakage of the elastic filaments by radiation leads to thick filament misalignment on stretch or activation. The M lines have been omitted for simplicity.

resting tension on stretch and also tends to keep the thick filaments centred within the sarcomere during generation of active force. These physiological functions are consistent with a sarcomere structure in which each end of the thick filament is mechanically linked to the nearest Z disk by elastic filaments made of titin and/or nebulin (Fig. 5). Such an organization is also consistent with much structural evidence, including the demonstration that gap filaments contain titin⁸; that titin-containing filaments span the region between thick filaments and Z lines²¹; and that small filaments emanating from the ends of thick filaments are morphologically similar to isolated titin²².

The model depicted in Fig. 5 also has important implications regarding the stability of the sarcomere during contraction. The sliding filament model predicts that the position of the thick filaments at the centre of each sarcomere is inherently unstable. This is because the amount of force on each half of the thick filament during contraction is proportional to the amount of overlap with thin filaments, so that any initial imbalance would be amplified as the thick filament is actively pulled toward one end of the sarcomere. The instability that is inherent in the conventional two-filament model of the sarcomere is reduced in several ways in models in which elastic filaments link the thick filaments to the Z-disks. First, as axial movement of thick filaments would lead to stretch of the elastic filaments, this movement would be resisted by the elastic filaments. Second, at sarcomere lengths where the elastic filaments are not stiff enough to completely prevent movement of the thick filaments (for example, near slack length), elastic filaments would be stretched as the thick filaments are actively pulled to one side of the sarcomere. These stretched elastic filaments would in

effect be transmitting active force to the Z disk, force which would otherwise never reach the ends of the sarcomere. Finally, on relaxation the elastic filaments would recentre the thick filaments in the sarcomere and thus optimize conditions for the start of another contraction.

Many authors have proposed sarcomere models similar to that shown in Fig. 5 on the basis of structural evidence^{4,8,23,24}. These models are now strongly supported by physiological evidence that titin and/or nebulin is essential for the maintenance of thick filament alignment and the efficient production of active and passive forces.

We thank Dr Kuan Wang for stimulating discussions.

Received 10 April; accepted 16 July 1986.

1. Carlsen, F., Knappeis, G. G. & Buchthal, F. *J. biophys. biochem. Cytol.* **11**, 95-117 (1961).
2. Huxley, A. F. & Peachey, L. D. *J. Physiol., Lond.* **156**, 150-165 (1961).
3. Locker, R. H. & Leet, N. G. *J. Ultrastruct. Res.* **52**, 64-75 (1975).
4. Magid, A., Ting-Beall, H. P., Carvell, M., Kontis, T. & Lucaveche, C. in *Contractile Mechanisms in Muscle* (eds Pollack, G. H. & Sugi, H.) 307-328 (Plenum, New York, 1984).
5. Sjostrand, F. S. *J. Ultrastruct. Res.* **7**, 225-246 (1962).
6. Wang, K. & Ramirez-Mitchell, R. *J. Cell Biol.* **96**, 562-570 (1983).
7. Maruyama, K. in *Muscle Contraction: Its Regulatory Mechanisms* (eds Ebashi, S., Maruyama, K. & Endo, M.) 485-496 (Springer, New York, 1980).
8. Wang, K. *Cell Muscle Motil.* **6**, 315-369 (1985).
9. Maruyama, K., Kimura, S., Yoshidomi, H., Sawada, H. & Kikuchi, M. *J. Biochem., Tokyo* **95**, 1423-1433 (1984).
10. Wang, K. & Williamson, C. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3254-3258 (1980).
11. Maruyama, K., Kimura, S., Ohashi, K. & Kuwano, Y. *J. Biochem., Tokyo* **89**, 701-709 (1981).
12. Wang, K., Wright, J. & Ramirez-Mitchell, R. *Biophys. J.* **47**, 349a (1985).
13. Wang, K., McClure, J. & Tu, A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3698-3702 (1979).
14. Kempner, E. S. & Haigler, H. T. in *Growth and Maturation Factors Vol. 3* (ed. Guroff, G.) 149-173 (Wiley, New York, 1985).
15. Kempner, E. S. & Haigler, H. T. *J. biol. Chem.* **257**, 13297-13299 (1982).
16. Magid, A. & Law, D. *J. Science* **230**, 1280-1282 (1985).
17. Podolsky, R. J. *J. Physiol., Lond.* **170**, 110-123 (1964).
18. Mannherz, H. G. & Goody, R. S. *A. Rev. Biochem.* **45**, 427-465 (1976).

19. Hanson, J., O'Brien, E. J. & Bennett, P. M. *J. molec. Biol.* **58**, 865-871 (1971).
20. Atwood, K. C. & Norman, A. *Proc. natn. Acad. Sci. U.S.A.* **35**, 696-709 (1949).
21. Maruyama, K., Yoshioka, T., Higuchi, H., Ohashi, K., Kimura, S. & Natori, R. *J. Cell Biol.* **101**, 2167-2172 (1985).
22. Trinick, J., Knight, P. & Whiting, A. *J. molec. Biol.* **180**, 331-356 (1984).
23. Higuchi, H. & Umazume, Y. *Biophys. J.* **48**, 137-147 (1985).
24. Maruyama, K., Sawada, H., Kimura, S., Ohashi, K., Higuchi, H. & Umazume, Y. *J. Cell Biol.* **99**, 1391-1397 (1984).
25. Wood, D. S., Zollman, J., Rueben, J. P. & Brandt, P. W. *Science* **187**, 1075-1076 (1975).
26. Somerville, L. L. & Wang, K. *Biochem. biophys. Res. Commun.* **102**, 53-58 (1981).
27. Thames, M. D., Teichholz, L. E. & Podolsky, R. J. *J. gen. Physiol.* **63**, 509-530 (1974).
28. Goldman, Y. E. & Simmons, R. M. *J. Physiol., Lond.* **350**, 497-518 (1984).
29. Hellam, D. C. & Podolsky, R. J. *J. Physiol., Lond.* **200**, 807-819 (1969).

Monoclonal antibodies to promote marrow engraftment and tissue graft tolerance

S. P. Cobbold, G. Martin, S. Qin & H. Waldmann

Department of Pathology, Cambridge University, Tennis Court Road, Cambridge CB2 1QP, UK

Allogeneic reactions are the major limitation to organ transplantation. These are manifested as rejection of the grafted tissue, and also, in the case of bone marrow transplantation (BMT), graft-versus-host disease (GVHD)¹. Recent methods of avoiding GVHD, by depleting T cells from donor marrow, have led to an increased incidence of marrow graft rejection²⁻⁵. Current recipient conditioning protocols involving drugs or irradiation cannot safely be increased, so alternatives must be found. Monoclonal antibodies can be used to control immune responses *in vivo*^{6,7}, and would be useful in this context if we could define and deplete the cells responsible for marrow rejection. We show here that elimination of residual L3T4⁺ and Lyt-2⁺ cells from mice receiving fully mismatched bone marrow abrogates rejection and promotes tolerance to donor-type skin grafts, even in sub-lethally irradiated recipients.

The rejection of allogeneic marrow is a serious limiting factor in the clinical application of bone marrow transplantation (particularly if there is no matched sibling donor³), as graft-versus-host disease can now be avoided by removing T cells from the donor marrow^{2,4,8-10}. It is clear that the recipients' 'resistance' to marrow grafting must be further reduced, but it is unlikely that current conditioning regimens could be increased as these are already near the limit of nonspecific toxicity. If marrow rejection could be prevented by a more specific and less toxic method, then BMT could be used more widely in, for example, the correction of inherited disorders, and perhaps to induce specific tolerance to other organ transplants. A number of mechanisms have been proposed to explain the resistance to marrow grafting, including natural killer (NK) cells^{11,12}, and antibody-dependent cell-mediated cytotoxicity (ADCC)¹³. Using mouse models of T-cell purged, mismatched BMT, we have recently shown that lethally irradiated recipients depleted of all T cells with monoclonal antibodies were much improved in long-term survival and chimaerism when compared to undepleted recipients¹⁴. It therefore became important to establish which T-cell subsets are responsible for rejecting marrow, and to determine whether monoclonal serotherapy to eliminate these subsets might supplement any sub-optimal conditioning of the transplant recipient.

Recipient CBA mice were given different doses of X irradiation, up to the lethal dose of 850 rad, either with or without pretreatment with monoclonal antibodies to deplete L3T4⁺ and Lyt-2⁺ T cells *in vivo*. The details of the protocol are shown in Fig. 1a. Acceptance or rejection of the T-cell depleted allogeneic marrow was determined by measuring both the level of chimaerism, and the degree of either unresponsiveness or priming to donor antigens as assessed by skin graft survival (Fig. 1b). Third

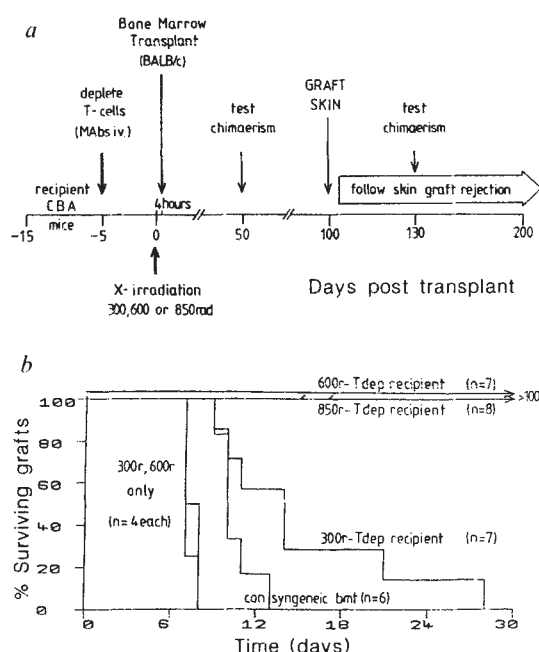


Fig. 1 Recipient T-cell depletion avoids marrow rejection and assists tolerance induction, even with sub-lethal irradiation. **a**, Experimental protocol. All animals were bred and maintained in the conventional animal facility of the Department of Pathology, Cambridge University. Groups of CBA/Ca mice (16-20 weeks old, 8-20 per group), some of which had received 50 µg monoclonal antibodies YTS 169.4+YTS 191.1 intravenously (sufficient to deplete all T cells by >90%)⁶, were given different doses of X irradiation (38 rad min⁻¹) on day zero. These were inoculated 4 h later with a mixture (10⁷ of each) of bone marrow and spleen cells (to be comparable to the blood-contaminated marrow used in human BMT) from BALB/c mice which had been adult thymectomized (ATX) and depleted of all T cells with antibodies to L3T4 and Lyt-2 (YTS 191.1 and YTS 169.4, 2 × 500 µg intravenously, at least 10 days previously^{6,14}). The recipient mice were given oral metronidazole (days -15 to -5), Terramycin (oxytetracycline, days -5 to 0), followed by pH 2 water (days 0 to 100) and then Terramycin again from day 100. The test for chimaerism is detailed in Fig. 2. Randomly selected 100-day survivors (4-8 per group) were simultaneously grafted with donor-type (BALB/c) and third party (C57/B10) tail skin grafts (~10 × 5 mm) in the same thoracic graft bed, separated by a syngeneic CBA/Ca graft. Dressings were removed on day 7 and the grafts observed for rejection for at least 100 days. Control animals were syngeneic mice transplanted with CBA/Ca bone marrow and equivalent in all other respects to the 850 rad allogeneic group. Overall survival for each group at 100 days was: 100% (300 rad), 88% (600 rad) and 40% (850 rad, T-depleted), and at 200 days: 100% (300 rad), 76% (600 rad) and 30% (850 rad, T-depleted). **b**, Responses to donor-type skin grafts. Analysis of graft survival was by the 'log-rank' method²⁶. Control syngeneic (CBA → CBA, 850 rad) rejection of BALB/c; median = 10 days. T-depleted donor only (300 or 600 rad with no recipient T depletion): median = 7 days, $P \leq 0.02$. T-depleted 300 rad recipients: median = 14 days, $P \leq 0.12$ versus controls. T-depleted 600 or 850 rad recipients: no rejection at 100+ days ($P \leq 0.001$), but two 850 rad animals died with grafts intact on days 14 and 17.

party skin was also grafted to establish the specificity of any unresponsive (tolerant) state. Recipients which had been T-cell depleted and lethally irradiated (850 rad) before marrow transplantation became specifically tolerant of donor-type antigens, as evidenced by complete skin graft acceptance. In contrast, similar recipients without T-cell depletion did not survive until day 100, as previously reported¹⁴. The majority of animals given 600 rad total body irradiation (TBI) survived, but with a remarkable difference between the T-cell depleted and untreated recipient groups: the former showed specific acceptance of donor-type skin grafts, while the latter rejected donor grafts



Fig. 2 Chimaerism in T-cell-depleted allogeneic bone marrow recipients. The levels of chimaerism were tested in the animals of Fig. 1 by immunofluorescence on whole blood smears using a rat monoclonal anti-mouse H-2 which binds to H-2K (recipient-type) but not H-2D²⁷. (This method has previously been shown to correlate well with full H-2 phenotyping by microcytotoxicity¹⁴.) Individual animals are shown as closed circles, while those with open circles had low blood white cell counts ($<5.0 \times 10^6 \text{ ml}^{-1}$ mostly neutrophils; normals were $6-8 \times 10^6 \text{ ml}^{-1}$ with approximately 60% lymphocytes). Where possible, 200+ mononuclear cells were counted when slides had $<10\%$ H-2K⁺ cells. Mean % recipient-type cells is indicated by the blocks. Polymorphs showed the same trend but were generally less positive for H-2K, even in control CBA mice (data not shown).

more rapidly than controls (that had received a syngeneic marrow graft). Third party grafts were rejected (data not shown) either at the normal rate or, in the group rejecting the marrow, at an accelerated rate (probably due to priming against cross-reactive transplantation antigens). A similar pattern was seen using animals irradiated with 300 rad, except that the antibody-treated recipients were now only partially unresponsive to donor-type skin grafts (Fig. 1b). The animals in this group were found to be mixed chimaeras, albeit predominantly recipient-type (Fig. 2), suggesting that this particular conditioning regime did not give the donor marrow sufficient advantage to compete effectively with an autologous recovery. In the case of T-cell-depleted mice given 600 rad or 850 rad, the peripheral blood cells reconstituted mainly to donor type (negative for recipient H-2; Fig. 2), and have remained so (>300 days), while the equivalent untreated (600 rad) group had an autologous reconstitution of the blood system. We conclude that monoclonal antibody depletion of recipient T cells prevents marrow graft rejection, leading instead to acceptance and specific tolerance of donor-type bone marrow and skin. Perhaps the most striking finding is that the autologous recovery expected after sub-lethal irradiation is reduced when the marrow is not rejected. This probably reflects a natural haemopoietic competition between the donor stem cells and the irradiated recipient cells.

What, then, is the mechanism of rejection, and which T cells are responsible? There are two major subsets of peripheral T cells in both man and rodents, which differ in the way they recognize foreign antigens¹⁵. The first of these responds to antigens in the context of class II (Ia) major histocompatibility complex (MHC) molecules¹⁶, and expresses the differentiation antigen L3T4 (ref. 17) (CD4 in man¹⁸). This subset can provide 'help' for other effector functions, and generate MHC class II-restricted effector cells. The second subset recognizes antigen in the context of class I MHC molecules¹⁹ (for example, H-2K, H-2D) and expresses the Lyt-2 differentiation antigen²⁰ (CD8 in man¹⁸). This subset is classically linked with the generation of cytotoxic T cells. To what extent, then, are each of these subsets capable of rejecting tissue grafts?

We have previously shown that L3T4⁺ T cells can rapidly reject either H-2 or minor antigen mismatched skin, while Lyt-2⁺

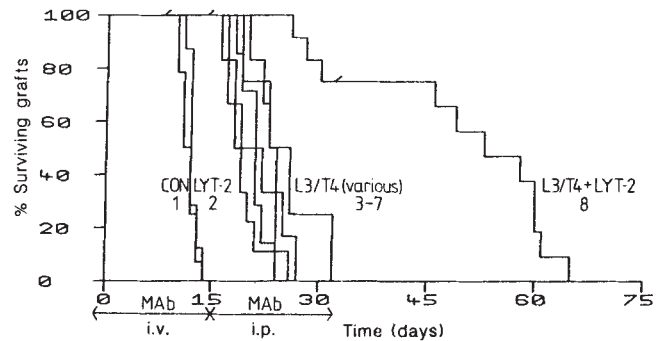


Fig. 3 Both L3T4⁺ and Lyt-2⁺ cells can reject allogeneic skin grafts. Adult CBA/Ca mice were grafted on day 0 with BALB/c tail skin on the lateral thoracic wall, by the method of Billingham *et al.*²⁸. The following schedules of monoclonal antibody treatment were used: (1) controls received no antibody ($n=14$); (2) YTS 169.4 (anti-Lyt-2)- approximately 500 μg intravenously on days -1, +2, +8, and intraperitoneally on day +12 ($n=8$, one animal died of anaphylaxis on day 8 with graft intact); (3) YTS 191.1 (anti-L3T4: epitope P), approximately 500 μg intravenously on days -1, +7, +12 and +19 ($n=6$); (4) YTA 3.1 (anti-L3T4: epitope Q), approximately 500 μg intravenously on days -1, +7, +12 and +19 ($n=6$); (5) YTS 191.1 + YTA 3.1 (250 μg of each) as in 3 and 4 ($n=9$); (6) YTS 191.1 + YTA 3.1 (25 μg of each) intravenously on days -1, +1 and three times per week thereafter, switching to intraperitoneal injection on day +15 until day +33 ($n=8$); (7) YTS 191.1 alone, doses as in 6 ($n=5$, one animal died on day 14 with graft intact); (8) YTS 191.1 + YTS 3.1 (10 μg of each synergistic anti-L3T4) and also YTS 169.4 + YTS 156.7 (10 μg of each synergistic anti-Lyt-2: S.Q. *et al.*, in preparation) as in 6 ($n=13$, one animal died with graft intact on day 32). All mice were given oxytetracycline (50 mg l^{-1} in the drinking water, Terramycin, Pfizer) and the antibody was 0.2 μm sterile filtered in PBS with penicillin (Crystapen; Glaxo; 500 U per injection) in groups 6, 7 and 8. Dressings were removed on day 10 and the grafts were observed daily, with the day of rejection recorded when no viable graft remained. All animals were shown to have no detectable ($<0.5\%$) cells positive for the depleted antigen on days 15 and 35 (control and anti-Lyt-2-treated day 15 only), by immunofluorescence on whole blood smears¹⁴ using the synergistic pairs and FITC-NORIG 7.16.2 (anti-rat IgG2b²⁹). Analysis of the survival curves by the log-rank method²⁶ gave the following P values of statistical significance: controls versus Lyt-2, not significant; Lyt-2 versus all L3T4, $P \leq 0.0001$; various L3T4 compared, $0.13 \leq P \leq 1.0$; L3T4 + Lyt-2 versus L3T4, $P \leq 0.0001$.

cells dominate the secondary graft response²¹. It was not clear, however, whether the Lyt-2⁺ cells could reject such grafts independently, or whether they needed to interact with L3T4⁺ 'helpers'. Recent data show that purified Lyt-2⁺ cells can respond *in vitro* to class I MHC differences without any help from L3T4⁺ cells²². The experiment in Fig. 3 shows that both subsets can indeed reject skin grafts relatively independently. Long-term survival of a fully mismatched skin graft was only obtained if both L3T4⁺ and Lyt-2⁺ cells were kept depleted. The fact that Lyt-2 depletion alone did not cause any delay confirms that L3T4⁺ cells can reject rapidly on their own. However, maintaining depletion of the L3T4⁺ cells, using a variety of aggressive protocols (including pairs of non-competing, rat IgG2b monoclonal antibodies which synergize for cell destruction *in vitro* and *in vivo*; refs 23, 24 and S.Q., unpublished) could not delay graft rejection beyond 35 days. The considerable extra suppression gained by eliminating both subsets demonstrates that the Lyt-2⁺ cells can also reject grafts without detectable L3T4⁺ cells, but only after some delay. It should be noted that antibody treatment alone did not impair the animals' ability to eventually reject their grafts, and was thus insufficient to impose a state of tolerance to this major antigenic challenge.

Bone marrow rejection, it seems, can also be elicited by either of the L3T4⁺ or Lyt-2⁺ subsets. In Fig. 4, it can be seen that,

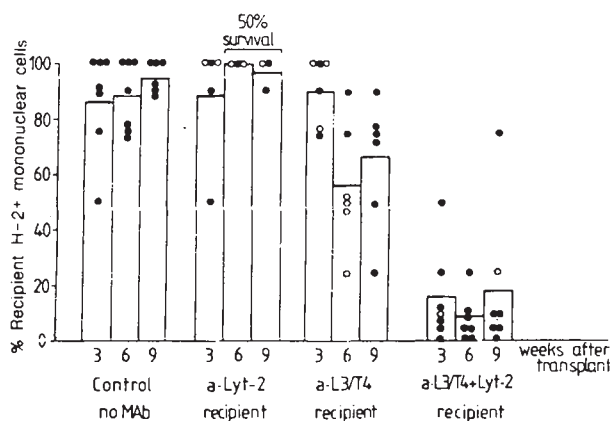


Fig. 4 Both L3T4⁺ and Lyt-2⁺ cells reject allogeneic marrow. Groups of CBA/Ca mice (14–18 weeks old) were treated according to the 600 rad protocol of Fig. 1, except that two new groups were included—separate recipient depletion of L3T4 (YTS 191.1, 50 µg intravenously on day -5) or Lyt-2 (YTS 169.4, 50 µg intravenously on day -5) with T-cell-depleted donor ATX BALB/c bone marrow and spleen cells, as before. Chimaerism was determined as described above, 3, 6 and 9 weeks post-transplant. Individual mice are shown as closed circles, while open circles depict a low white cell count (mostly polymorphs) as in Fig. 3 (except anti-Lyt-2 group on week 3 were mainly lymphocytes of recipient type). Mean % recipient-type cells is indicated by the blocks.

after sub-lethal irradiation (600 rad TBI), histoincompatible marrow transplants were only accepted by mice depleted of both the T-cell subsets. Recipients depleted of just one subset rejected the donor marrow and reconstituted instead with autologous blood cells, as did untreated controls. This shows that both subsets have an independent capacity to reject allogeneic marrow. None of the animals depleted of Lyt-2 cells had any evidence of significant donor cells in the blood, and 3 out of 6 mice died before day 28, suggesting some inhibition of autologous reconstitution. The L3T4⁺ cells seemed more effective than Lyt-2⁺ cells for marrow rejection, because there was evidence of partial chimaerism in the L3T4-depleted group. However, these mice rejected donor-type skin grafts after only 8 days (median: data not shown), indicating that they had become primed to the donor marrow antigens. This is in contrast to the tolerance-inducing effects of anti-L3T4 treatment as recently described for antibody responses to soluble proteins²⁵. It is likely that tolerance to donor transplantation antigens is due to the continuous presence of donor haematopoietic cells, which is only possible after acute marrow rejection has been avoided by removing the recipients' T cells (which can be achieved with either anti-L3T4+Lyt-2 as described above, or with anti-Thy-1¹⁴).

In practical terms, it is possible that the ability to improve the conditioning of bone marrow transplant recipients with appropriate monoclonal antibodies will be directly applicable to man. This would simplify mismatched marrow transplantation, and perhaps allow nonspecific ablative conditioning to be reduced to a level where BMT might be used to induce tolerance to other donor organ grafts.

This work was supported by the UK MRC.

Received 19 March; accepted 11 June 1986.

- Ortho Multicentre Transplant Study Group *New Engl. J. Med.* **313**, 337–342 (1985).
- Rodt, H., Thierfelder, S. & Eulitz, M. *Eur. J. Immun.* **4**, 25–29 (1974).
- Trentin, J. J. & Judd, K. P. *Transplant Proc.* **5**, 865–868 (1973).
- Prentice, H. G. *et al. Lancet* **i**, 472–476 (1984).
- Cudkovic, G. & Bennett, M. *J. exp. Med.* **134**, 1513 (1971).
- Waterfall, M., Rayfield, L. S. & Brent, L. *Nature* **311**, 663–665 (1984).
- Warner, J. F. & Dennert, G. *J. exp. Med.* **161**, 563–576 (1985).
- Cobbold, S., Martin, G. & Waldmann, H. *Transplantation* **42** (in the press).
- Sprent, J. & Schaefer, M. *J. exp. Med.* **161**, 2068 (1985).
- Katz, D. H., Hamaoka, T. & Benacerraf, B. *J. exp. Med.* **137**, 1405–1418 (1973).
- Dialynas, D. P. *et al. Immun. Rev.* **74**, 29–56 (1983).
- Bernard, A., Boumsell, L. & Hill, C. in *Leucocyte Typing: Human Leucocyte Differentiation Antigens Detected by Monoclonal Antibodies Specification-Classification-Nomenclature* (eds Bernard, A. *et al.*) (Springer, New York, 1984).
- Zinkernagel, R. M. & Doherty, P. C. *Nature* **251**, 547–548 (1974).
- MacDonald, H. R., Thiernes, N. & Cerottini, J. C. *J. Immun.* **126**, 1671 (1981).
- Cobbold, S. P. & Waldmann, H. *Transplantation* **41**, 634 (1986).
- Mizuochi, T. *et al. J. exp. Med.* **162**, 427–443 (1985).
- Hughes-Jones, N. C., Gorick, B. D., Miller, N. G. A. & Howard, J. C. *Eur. J. Immun.* **14**, 974–978 (1984).
- Bindon, C. I., Hale, G., Clark, M. R. & Waldmann, H. *Transplantation* **40**, 538–544 (1985).
- Benjamin, R. & Waldmann, H. *Nature* **320**, 449–451 (1986).
- Peto, R. *et al. Br. J. Cancer* **35**, 1–39 (1977).
- Koch, S., Koch, N., Robinson, P. & Hammerling, G. *Transplantation* **36**, 177–180 (1983).
- Billingham, R. E., Brent, L. & Medawar, P. B. *Proc. R. Soc. B143*, 58 (1954).
- Aqel, N. M., Clark, M., Cobbold, S. P. & Waldmann, H. *J. Immun. Meth.* **69**, 207–214 (1984).

Isolation of cell lines possessing functional and serological properties resembling those of thymocyte precursors

Leslie O. Goodwin, Antonio J. D. Rocha & Ross S. Basch

Department of Pathology and Kaplan Cancer Center, New York University Medical Center, 550 First Avenue, New York, New York 10016, USA

Thymocytes develop from a committed haematopoietic progenitor, referred to as a prothymocyte^{1–4}. They are uniquely capable of migrating to and restoring the thymus of a lethally irradiated host, a property which has been exploited as a specific assay for these cells^{4–6}. Like other committed haematopoietic progenitors, prothymocytes are found only in small numbers in even the richest sources (0.05–1.0% of the nucleated cells in bone marrow). Purification has proved difficult both in terms of finding a suitable starting material and in the degree of enrichment achieved^{7,8}. We now report the isolation of cloned lines of cells with some of the serological and functional properties of prothymocytes. One of these lines has been in continuous culture for almost 2 years. When injected into irradiated recipients, cells from this line migrate to the thymus and there develop into cells which resemble normal cortical thymocytes.

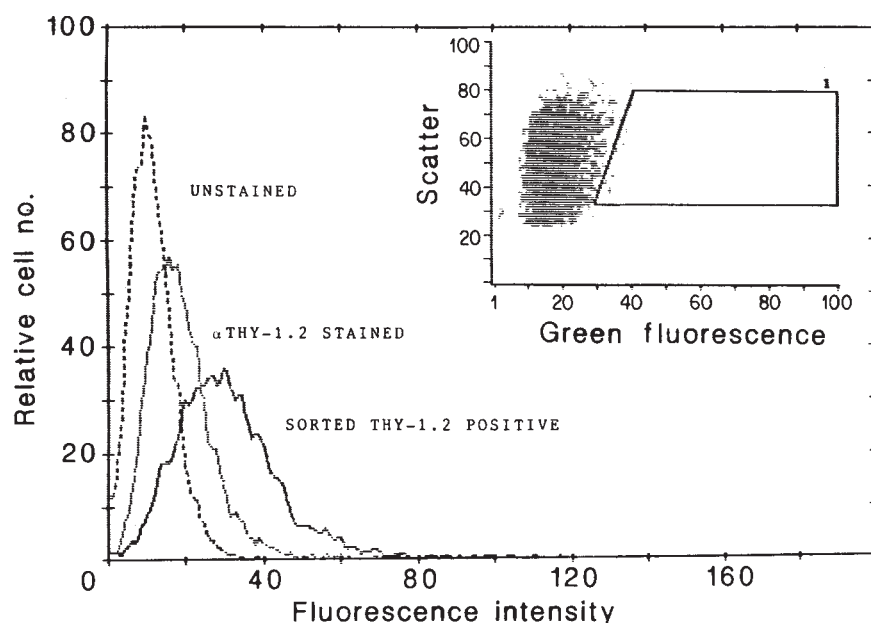
Attempts have been made to culture thymocyte precursors from a variety of haematopoietic tissues. Our first isolate was obtained from a mixed leukocyte culture (MLC) in which spleen cells from BALB/c mice were cultured with irradiated (2,000 rad) C57BL/6 spleen cells in Iscove's modified Dulbecco's minimal essential medium supplemented with 10% (v/v) conditioned medium from rat spleen cells stimulated by concanavalin A (Rat T-cell monoclonal; Collaborative Research, Lexington) and 10% fetal bovine serum. One week after the initial burst of proliferation of allogeneic T cells had subsided, clusters of large blasts appeared. These were collected by aspiration and transferred into fresh medium as above. When the cultures reached a density of 500,000 cells ml⁻¹ they were diluted 1:5. After 5 weeks, they were stained for surface Thy-1.2 antigen and the apparently negative population (95% of the total) was isolated by sorting. This population was cloned using a single-cell deposition device. The initial cloning efficiency was 5–10%. The cells were recloned twice and one of these clones (now designated pT-D8) was selected for further study.

After 6 months in culture, supplementation with conditioned medium became unnecessary. Recent isolates of similar clones

- Vriesendorp, H. M. in *Bone Marrow Transplantation. Biological Mechanisms and Clinical Practice* (eds van Bekkum, D. W. & Lowenberg, B.) 73–146 (Dekker, New York, 1985).
- Waldmann, H. *et al. Lancet* **ii**, 483–486 (1984).
- Sondel, P. M. *et al. Transplant Proc.* **17**, 460–461 (1985).
- O'Reilly, R. J. *et al. Transplant Proc.* **17**, 455–459 (1985).
- Martin, P. J., Hansen, J. A., Storb, R. & Thomas, E. D. *Transplant Proc.* **17**, 486–487 (1985).
- Cobbold, S. P., Jayasuriya, A., Nash, A., Prospero, T. D. & Waldmann, H. *Nature* **312**, 548–551 (1984).

Fig. 1 Histogram of the immunofluorescence of pT-D8 stained with fluoresceinated anti-Thy-1.2 monoclonal antibody and a re-analysis of the cells sorted as Thy-1⁺. The insert shows the window used to sort the Thy-1⁺ pT-D8 cells.

Methods. Immunofluorescence was performed as described previously³⁰. For indirect immunofluorescence, cells were first treated with the primary reagent and, after washing, stained with either a fluoresceinated anti-immunoglobulin or phycoerythrin-conjugated avidin. When only one fluor was used, the cells were also stained with propidium iodide to identify dead cells. Cell-surface fluorescence was quantified using an Ortho 50H cytofluorograph equipped with a 5-W argon laser operating at 200 mW at 488 nm. Phycoerythrin fluorescence was detected by using a dichroic mirror reflecting light of wavelength longer than 560 nm. Fluorescein emission was measured by filtering the light passed by the dichroic through a narrow-band green filter (530 ± 10 nm).



have lost their dependence on exogenous growth factors 6–8 weeks after removal from the MLC. Allogeneic stimulation is not necessary for the development of these clones. Similar lines have been developed by initiating the cultures in medium conditioned by concanavalin A-stimulated spleen cells and then maintaining the cells in medium containing purified interleukin-

Table 1 Surface markers of pT-D8 compared with normal thymocytes and pT-D8 progeny

Antibody*	2nd reagent	pT-D8 (mean fluorescence intensity)	Thymocytes	D8 Progeny*
Thy-1.2 (biot.)	Fl-avidin	16	1,543	†
Lyt1 (biot.)	Fl-avidin	4	154	162
Lyt2 (biot.)	Fl-avidin	6	483	454
L3T4 (biot.)	Fl-avidin	153	237	216
Thy-1.2	Fl-GaMIG	0	501	ND
Thy-1.1	Fl-GaMIG	12	15	ND
Ly 5	Fl-GaMIG	1,064	495	ND
Qa-2	Fl-GaMIG	0	15	ND
T1	Fl-GaMIG	0	44	ND

Net mean fluorescence intensity is given in arbitrary units. Values obtained for cells stained with the same second reagent can be compared but comparisons between groups are not valid. ND, not done. Monoclonal (mouse IgM) anti-Thy-1.1 and -Thy-1.2 antibodies were purchased from NEN. A biotinylated (biot.) anti-Thy-1.2 monoclonal antibody (clone 30H12, rat IgG2b; ref. 25) was purchased from Becton Dickinson as were biotin-conjugated and fluoresceinated (Fl) anti-Lyt1 (clone 53-7.3, rat IgG2a; ref. 25) and anti-Lyt2 (clone 53-6.7, rat IgG2a; ref. 25). Anti-L3T4 was purified from the culture fluid of hybridoma clone GK1.5 (rat IgG2b (ref. 20) given to us by Dr Frank Fitch. Anti-Ly5 (clone 103-2ES11p53, mouse IgG; ref. 26), anti-Ia (clone M5114, rat IgG2b; ref. 27) and anti-T1 (clone m4, mouse IgG2b; ref. 28) were provided by Dr F. W. Shen. Anti-Qa-2 (D3.262 pool 10, mouse IgM; ref. 29) was a gift of Dr L. Flaherty. Goat anti-mouse IgG (fluorescein conjugated; Fl-Ga MIG) was purchased from Cappel Laboratories and affinity-purified fluorescein-conjugated F(ab)₂ mouse anti-rat IgG reagent was purchased from Jackson Immunolabs.

* Cells were recovered from the thymus of (AKR/J × BALB/c)F₁ mice 14 days after irradiation and repopulation with 5 × 10⁶ pT-D8 cells, and purified by sorting for Thy-1.2⁺, Thy-1.1⁺ thymocytes as described in the text.

† These cells were stained with biotinylated Thy-1.2 and phycoerythrin-avidin and sorted to obtain a homogeneous population of pT-D8 progeny. Since they have already been stained with a Thy-1.2 reagent, they cannot be retested with the same reagent.

3 (ref. 9). A line has also been developed using spleen cells from a BALB/c nude (nu/nu) mouse. Cloning efficiencies of the factor-independent clones are 60–70%.

When stained with monoclonal antibodies to Thy-1.2, most of the pT-D8 population stains dimly (Table 1, Fig. 1). A small proportion (0.05–0.1%) of the cells stain as brightly as normal peripheral T cells. This staining is not an artefact produced by dead cells because the brightly staining cells exclude propidium iodide. Similar Thy-1⁺ cells have been found in all of the isolates examined. If pT-D8 are stained with anti-Thy-1 reagents and sorted on the basis of their fluorescence intensity, a population with intermediate staining properties can be isolated (Fig. 1). In culture, these cells lose Thy-1.2 expression, and after 1 week cannot be distinguished from the unselected parent line. After 3 weeks, small numbers of brightly staining, Thy-1.2⁺ cells reappear in the cultures of Thy-1⁺ cells. Attempts to culture these brightly staining cells have been uniformly unsuccessful.

Cells from clone pT-D8 have been grown in mass culture and then tested for the presence of a variety of thymic differentiation antigens (Table 1). The staining of normal thymocytes is shown for purposes of comparison. The intensity of the staining of pT-D8 with anti-Thy-1.2 antibody is 1–2% of that of normal thymocytes, but the surface antigen Ly 5 is present in greater amounts than found on thymocytes. The antigen L3T4 is expressed by pT-D8 at a relatively high level (approximately two-thirds of that found on BALB/c thymocytes). Trace quantities of Lyt2 were found on one line (which has since been lost) but none of the other antigens tested are expressed on these cells. L3T4 expression has been detected on all of the lines including that derived from a nude mouse.

If pT-D8 cells are injected intravenously into irradiated syngeneic mice, the repopulation of the thymus occurs more rapidly than in uninjected animals (Fig. 2). Increased cell division in mice injected with pT-D8 can be detected as early as 9 days after the injection of as few as 2 × 10⁵ cells. By contrast, 2 × 10⁶ bone marrow cells produce increased intrathymic proliferation only after 12–15 days. The recoveries of cells from the thymus early in repopulation are related to the number of pT-D8 cells injected. Thus, on day 9 an uninjected animal yielded 0.52 × 10⁵ thymocytes while one receiving 2 × 10⁶ bone marrow cells yielded 1.2 × 10⁶ cells; 3.9 × 10⁶, 1.7 × 10⁶ and 1.4 × 10⁶ cells were recovered from the thymuses of animal receiving 200,000, 2 × 10⁶ and 2 × 10⁷ pT-D8 cells, respectively.

To eliminate the possibility that the clones promote endogenous repopulation rather than actually serving as

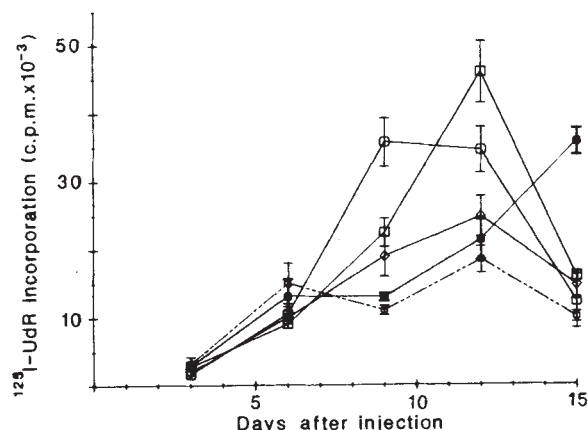


Fig. 2 Incorporation of ^{125}I -labelled deoxyuridine (UdR) by regenerating thymus after X-ray irradiation. —○—, No cells; —●—, 2×10^6 bone marrow; —□—, 2×10^5 pT-D8; —■—, 2×10^6 pT-D8.

Methods. BALB/c mice were irradiated with 790 rad and bone marrow or pT-D8 cells were injected 2 h later. At the indicated times, ^{125}I -UdR (10 μCi per mouse in phosphate-buffered saline) was injected intravenously and the thymus removed 4 h later. Three to five animals were used for each point and the results shown are the means of the individual counts. (The error bars indicate the standard deviation of the results.) The failure of pT-D8 to produce prolonged expansion of the thymocyte compartment as indicated by the decline in ^{125}I -UdR incorporation found on day 15 could indicate some abnormality in their developmental potential. However, normal bone marrow prothymocytes, injected intrathymically, produce only a single wave of repopulation³¹. Furthermore, 2 weeks after irradiation, most animals that have not been protected with bone marrow cells are moribund and cachexia is known to interfere with thymic repopulation².

thymocyte precursors, we substituted (AKR/J $\times$ BALB/c)F₁ mice as recipients and measured repopulation by immunofluorescence. Regenerating thymocytes of endogenous origin in F₁ mice express both Thy-1.1 and Thy-1.2, and therefore are doubly labelled in a two-colour immunofluorescence assay. The progeny of pT-D8 should express only Thy-1.2. The results of such an experiment are shown in Fig. 3. Most of the Thy-1⁺

regenerating cells are of pT-D8 origin. In another experiment in which individual mice were tested, the proportion of donor cells (that is, of pT-D8 origin) in the regenerating thymuses varied between 2.3% and 68.0%.

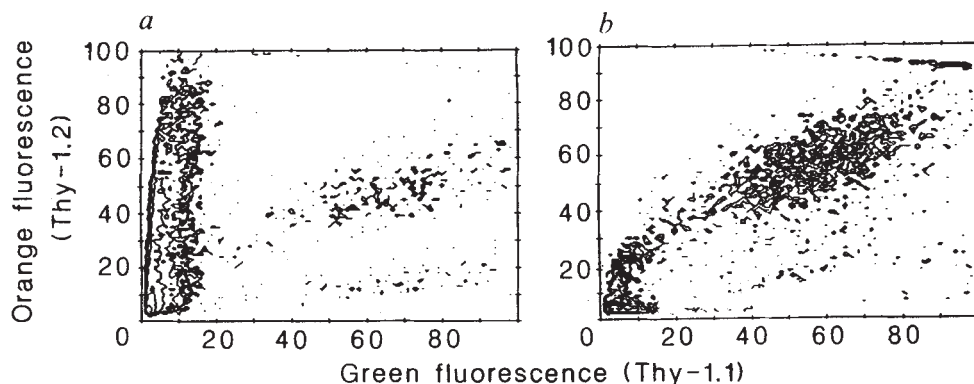
An essentially homogeneous population of cells bearing only Thy-1.2 and stained only with phycoerythrin can be obtained by sorting the cells shown in Fig. 3a. These cells can then be stained with fluoresceinated reagents in order to measure the expression of other differentiation antigens. Essentially all of the sorted cells (>90%) are positive for Lyt1, Lyt2 and L3T4; no evidence of bimodal staining was seen. Mean fluorescence is shown in the last column of Table 1. On the basis of these surface markers, the progeny of pT-D8 cannot be distinguished from normal cortical thymocytes. The presence of small numbers of cells expressing the phenotypes of mature T cells (Lyt2⁺, L3T4⁺ or Lyt2⁺, L3T4⁺) cannot be excluded because of the assay system. The cells are sorted on the basis of being stained with phycoerythrin. They are then stained with a fluoresceinated monoclonal antibody. A third fluorochrome would be required to detect cells expressing both Lyt2 and L3T4. Attempts to grow the sorted cells either in regular medium or the conditioned medium used to isolate the parent line have been unsuccessful.

To determine whether the cultured cells had the capacity to form tumours, 10^7 cells were injected (subcutaneously, intraperitoneally or intravenously) into groups of syngeneic mice (20 mice per group, half of which had been irradiated with 240 rad), and all animals were observed for 180 days. None of the 60 BALB/c mice developed tumours.

The degree of differentiation of thymocyte precursors at the time they become distinct from their multipotential ancestors or at the time of their entry into the thymus is not clear. Neither the timing of clonotypic differentiation (as evidenced by rearrangement of the genes encoding the antigen-specific T-cell receptor) nor the number of different precursors required to produce the variety of separate effector cells has been established^{7,10-15}. The precursors are 'null' cells, that is, they lack surface immunoglobulin and sufficient Thy-1 antigen to be detected by conventional alloantisera, and they are usually found in bone marrow, spleen and fetal liver. The most convincing criterion for prothymocyte function is the capacity to repopulate the thymus of an animal whose endogenous precursor population has been depleted by irradiation¹⁶. To achieve this, the cells must migrate to the organ, differentiate and proliferate. All of these tests are met by pT-D8. No other cells, maintained in continuous culture, meet these criteria.

Fig. 3 Two-colour immunofluorescence histogram of cells, obtained from the thymuses of (AKR/J $\times$ BALB/c)F₁ mice, regenerating after X-ray irradiation. The mice had received 790 rad 15 days previously. *a*, Mice injected with 10^7 pT-D8 cells; *b*, control irradiated mice. Ordinate: the phycoerythrin-avidin staining of Thy-1.2 (biotin)+ cells. Abscissa: fluoresceinated goat anti-mouse immunoglobulin staining of Thy-1.1+ cells.

Methods. At least 58% of the cells in *a* are of pT-D8 (that is, BALB/c) origin (78% of the thymocytes found in these animals were Thy-1⁺ and 75% of these stained only with the anti-Thy-1.2 reagent). In control F₁ mice, which had been irradiated but did not receive pT-D8, the regenerating thymocytes are Thy-1⁺ (87%) but 92% of these stain with both anti-Thy-1.1 and anti-Thy-1.2. Virtually none of the cells stain for Thy-1.2 only. For two-colour immunofluorescence, dead cells were excluded by gating on right-angle scatter. The cells were first stained with anti-Thy-1.1 and F1-anti-mouse immunoglobulin and then washed in phosphate-buffered saline containing mouse serum to prevent the bound anti-immunoglobulin from subsequently crossreacting with the rat anti-Thy-1.2 (biotin). The cells were then stained with the biotinylated anti-Thy-1.2 and phycoerythrin-avidin and examined in the cytofluorograph. In one experiment, the phycoerythrin-stained, Thy-1.2⁺, Thy-1.1⁺ cells were sorted and then stained with F1-Lyt1, F1-Lyt2 or biotinylated L3T4 and F1-avidin.



The surface phenotype of pT-D8 is consistent with its characterization as a thymocyte precursor. The presence of low levels of Thy-1 on these cells is well known and Ly 5 is widely distributed in the haematopoietic system¹⁷⁻¹⁹. The high level of L3T4 is surprising in view of the widely held model for thymocyte development, in which all Thy-1⁺, L3T4⁺ cells are assessed to develop intrathymically from Thy-1⁺, L3T4⁻ precursors^{10,11}. However, L3T4 is not restricted to thymocytes and mature T cells; approximately half the cells in the marrow express this antigen²⁰. Furthermore, direct measurements on bone marrow cells indicate that both L3T4⁺ and L3T4⁻ thymocyte precursors exist in normal mice (G. Frederickson, personal communication). While the presence of L3T4 on the clones described here does not establish the presence of the antigen on all thymocyte precursors, it does suggest its presence on some.

Could these cells be an aberrant form of post-thymic cell? We cannot exclude this possibility absolutely, but it seems unlikely. First, similar cells can be isolated from nude mice. While some T-cell maturation occurs in these animals²¹, it is certainly not under thymic influence, and to consider cells isolated from these mice as post-thymic seems unreasonable. It is also difficult to imagine the process of differentiation that would be required to produce a post-thymic cell that had lost Thy-1, Lyt1 and Lyt2 but retained the ability to home to the thymus and re-express, these antigens in the thymic environment.

pT-D8 do not induce tumours in normal mice and appear to migrate normally. In this limited sense they are 'normal' cells, but it is doubtful whether any cloned line with unrestricted proliferative potential is truly normal²². 'Immortalization' seems to be an early step in the transformation process which leads to malignancy^{23,24}. If this is true, this step must be a reversible one because pT-D8 progeny lose their unlimited, factor-independent growth potential during their thymic passage.

We thank Drs Joan W. Berman and Gay Fredrickson for helpful suggestions and discussions, and Allen Tate III for technical assistance. The Ortho 50H cytofluorograph was purchased and is operated by the Kaplan Cancer Center as part of its instrumentation facility, and we thank Ms Betty Goon for operating the instrument. This work is supported by NIH grants CA 24472 and CA 33104 and by the Lila Motley Cancer Fund.

Received 28 February; accepted 5 June 1986.

Growth restoration of insulin-deficient diabetic rats by recombinant human insulin-like growth factor I

Erika Scheiwiller*, Hans-Peter Guler*, Jim Merryweather†, Carl Scandella†, Walter Maerki‡, Jürgen Zapf* & E. Rudolf Froesch*

* Metabolic Unit, University Hospital, CH-8091 Zurich, Switzerland

† Chiron Corporation, Emeryville, California 94608, USA

‡ Ciba-Geigy, 4002 Basel, Switzerland

Insulin-like growth factor I (IGF-I) and insulin stem from a common precursor, are structural homologues, act through similar receptors and elicit insulin-like and growth-promoting effects *in vitro* and *in vivo*^{1,2}. Serum IGF-I levels are controlled by growth hormone¹, insulin³⁻⁶ and nutrition⁶⁻¹⁰. Insulin-deficient growth-arrested diabetic animals have reduced serum IGF-I levels which are restored towards normal by insulin but not by growth-hormone treatment³⁻⁶. Here we show that normal growth of diabetic rats is restored by infusion of recombinant human (rh)IGF-I without normalization of the blood sugar level and that insulin acts via an increase of IGF-I synthesis on growth of diabetic rats. We describe a new mechanism of endocrine control of growth in which IGF-I is the major stimulator at the cellular level. Growth hormone and insulin act mainly by modulating the hepatic synthesis of IGF-I.

Male Tif RAI rats (120-130 g) were used in all experiments. Diabetes was induced by streptozotocin which was freshly dissolved in 0.9% NaCl (pH 3.5 adjusted with citric acid) and injected into the tail vein in a dose of 110 mg per kg body weight. The rats had not received any food for 24 h and were under sedation with Valium (diazepam, 10 mg per kg body weight). Rats were used only if their body weight remained constant during the second and third week of diabetes. During treatment, the rats had free access to food (Altromin, Lage) and drinking water, and were kept on a 12-h light and dark cycle. The non-diabetic rats were age-matched and had reached a body weight of 260 ± 10 g at the time of killing. Insulin and growth hormone were administered in 0.9% NaCl, rhIGF-I in 0.1 M acetic acid. The hormones were administered for 6 days by subcutaneously implanted Alzet miniosmotic pumps (model 2001, Alza).

The body weight of the rats was recorded daily at 8.30 a.m. After 6 days of infusion the rats were anaesthetized with pentobarbital, bled by aortic puncture and blood glucose immediately determined in a glucose analyser. The two lowest pairs of ribs were removed and ³H-thymidine incorporation into DNA of costal cartilage was determined as described previously¹¹. The tibial epiphyseal width was measured according to Greenspan *et al.*¹². Aliquots of serum of each group were pooled and 0.5 ml gel-filtered over Sephadex G-50 in 1 M acetic acid according to the method of Schlumpf *et al.*¹³ to dissociate and separate IGF-I from the IGF carrier proteins. IGF-I (rhIGF-I and rat IGF-I) levels were measured by radioimmunoassay using an antiserum against human IGF-I that cross-reacts with rat IGF-I. In this assay, rhIGF-I and rat IGF-I yield parallel displacement curves with human ¹²⁵I-labelled IGF-I as a tracer.

In healthy rats, the mean serum IGF-I level is 155 ng ml⁻¹ whereas in diabetic rats it is reduced to 22 ng ml⁻¹. Administration of a large dose of human growth hormone (hGH, 400 mU per day) has no effect on IGF-I levels of diabetic rats and does not stimulate any of the growth parameters. 0.5 IU insulin per day and 300 µg IGF-I per day stimulate all measured growth indices to the same extent (Figs 1, 2). The insulin-treated rats have a mean endogenous IGF-I level of 66 ng ml⁻¹; those receiving 300 µg of rhIGF-I per day have a serum IGF-I concentration of 280 ng ml⁻¹. Because rat IGF-I is at least three

1. Ford, C. E. *et al.* *Ann. N.Y. Acad. Sci.* **129**, 283-296 (1966).
2. Basch, R. S. & Kadish, J. L. *J. exp. Med.* **145**, 405-419 (1977).
3. Komuro, K. & Boyse, E. A. *J. exp. Med.* **138**, 479-482 (1973).
4. Kadish, J. L. & Basch, R. A. *J. exp. Med.* **143**, 1082-1099 (1976).
5. Boersma, W. *et al.* *Cell Tissue Kinet.* **14**, 179-196 (1981).
6. Lepault, F. & Weissman, I. L. *Nature* **293**, 151-154 (1981).
7. Basch, R. S., Kadish, J. L. & Goldstein, G. *J. exp. Med.* **147**, 1843-1848 (1978).
8. Greiner, D. L., Goldschneider, I. & Barton, R. W. *J. exp. Med.* **156**, 1448-1460 (1982).
9. Ihle, J., Keller, N., Henderson, L., Klein, C. & Palaszynski, E. *J. Immun.* **129**, 2431 (1982).
10. Mathieson, B. J. & Fowlkes, B. J. *Immun. Rev.* **82**, 141-173 (1984).
11. Scollay, R. & Shortman, K. *Thymus* **5**, 245-295 (1983).
12. Goldschneider, I. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **78**, 2469-2473 (1981).
13. Born, W. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **82**, 2925-2929 (1985).
14. Chervenak, R., Moorehead, J. W. & Cohen, J. H. *J. Immun.* **134**, 695-698 (1985).
15. Katsura, Y. *et al.* *J. Immun.* **135**, 3021-3027 (1985).
16. Lesley, J., Hyman, R. & Schulte, R. *Cell. Immun.* **91**, 397-403 (1985).
17. Loo, F. & Roelants, G. E. *Ann. N.Y. Acad. Sci.* **254**, 226-240 (1975).
18. Basch, R. S. & Berman, J. W. *Eur. J. Immun.* **12**, 359-364 (1982).
19. Tung, J. S. *et al.* *Immunogenetics* **14**, 101-106 (1981).
20. Dialynas, D. P. *et al.* *Immun. Rev.* **74**, 29-56 (1983).
21. MacDonald, H. R. *Expl. cell. Biol.* **52**, 2 (1984).
22. Eftros, R. B. & Walford, R. L. *Hum. Immun.* **9**, 49-56 (1984).
23. Land, H., Parada, L. F. & Weinberg, R. A. *Nature* **304**, 596-602 (1983).
24. Cairns, J. & Logan, J. *Nature* **304**, 582-583 (1983).
25. Ledbetter, J. A. & Herzenberg, L. A. *Immun. Rev.* **47**, 63-90 (1979).
26. Shen, F. W. in *Monoclonal Antibodies and T-cell Hybridomas; Perspective and Technical Advances* (eds Hammerling, G., Hammerling, V. & Kearney, J. F.) 25-31 (1981).
27. Bhattacharya, A., Dorf, M. E. & Springer, T. A. *J. Immun.* **127**, 2488-2492 (1981).
28. Bernstein, I. D., Tam, M. R. & Nowinski, R. C. *Science* **207**, 68-71 (1980).
29. Lynes, M. A., Tonkonogy, S. & Flaherty, L. J. *Immunology* **129**, 928-930 (1982).
30. Berman, J. W. & Basch, R. S. *J. immun. Meth.* **36**, 335-338 (1980).
31. Goldschneider, I., Komschlies, K. L. & Greiner, D. L. *J. exp. Med.* **163**, 1 (1986).

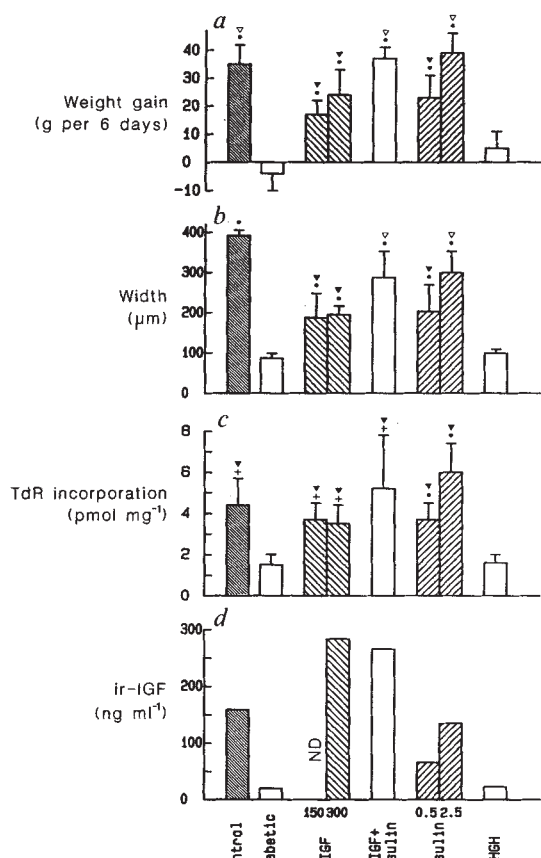


Fig. 1 Stimulation of growth indices in diabetic rats by insulin, IGF-I and hGH. *a*, Gain in body weight; *b*, tibial epiphyseal width; *c*, thymidine (TdR) incorporation; *d*, immunoreactive IGF-I concentration. Hormones were administered over 6 days by osmotic minipumps in the following daily amounts: insulin, 0.5 IU ($n=5$) and 2.5 IU ($n=6$); hGH, 400 mU ($n=5$); rhIGF-I, 150 µg ($n=4$) and 300 µg ($n=11$); a combination of 300 µg rhIGF-I and 0.5 IU insulin ($n=6$). Controls, non-diabetic rats of the same age. Bars represent means; brackets, standard deviations. Student's *t*-test was used for statistical analysis. All values in each panel were tested against the value of the untreated diabetic rats: •, $P<0.001$; +, $P<0.01$. All values of each treatment group were also tested against each other: ▽, no significant difference against values designated by ▽; ▽, no significant differences against values designated by ▽; ▽ versus ▽, $P<0.05-0.001$. In the radioimmunoassay for IGF-I, human ¹²⁵I-IGF-I served as a tracer and human IGF-I for standard dilutions. Rat IGF-I levels are, therefore, equivalents of human IGF-I and should not be considered as absolute values. It is impossible to distinguish between endogenous rat IGF-I and infused rhIGF-I. ND, not determined.

times more active on rat osteoblasts *in vitro* than human IGF-I¹⁴ the growth potential of endogenous rat IGF-I and rhIGF-I levels in these two groups of rats is comparable (Figs 1, 2).

The combination of 300 µg IGF-I together with 0.5 IU insulin per day stimulates growth to the same extent as 2.5 IU insulin alone. However, blood sugar is normalized only by the high dose of insulin (9.8 ± 6.2 mmol l⁻¹ versus 30.8 ± 5.2 in untreated diabetic rats) and remains elevated in the group receiving the small dose of insulin plus IGF-I (22.7 ± 2.8 mmol l⁻¹) or 300 µg IGF-I (23.2 ± 3.3). Thus, IGF-I administered to diabetic rats mimics the effect of insulin with respect to weight gain and growth although hyperglycaemia persists. In agreement with these findings the administration of pure IGF-I to hypophysectomized rats in doses which stimulate growth and mimic the effects of hGH do not influence blood glucose^{11,15}. Only large

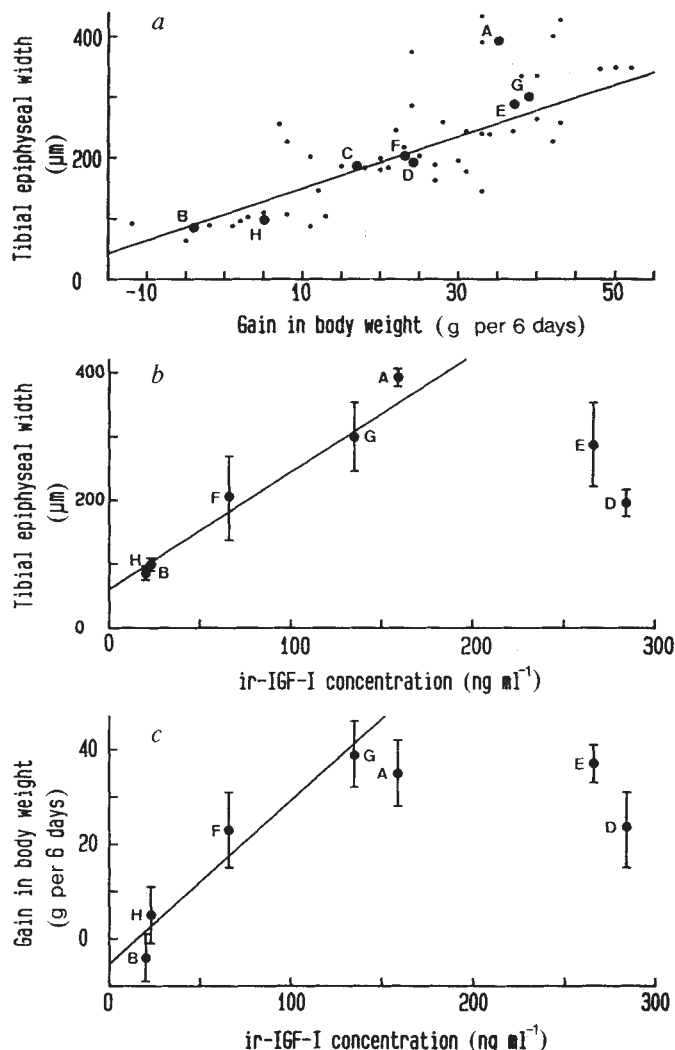


Fig. 2 Correlation between growth indices (*a*) and IGF-I concentrations (*b*, *c*) in untreated and treated streptozotocin diabetic rats. Treatment with hormones was as described in Fig. 1. A, non-diabetic rats of the same age; B, untreated; C, rhIGF-I, 150 µg; D, rhIGF-I, 300 µg; E, rhIGF-I, 300 µg plus insulin, 0.5 IU; F, insulin, 0.5 IU; G, insulin, 2.5 IU; H, growth hormone. Group A was not used for the calculation of the regression *r* (*a-c*); groups E and D were not used in *b* and *c*. $r=0.83$ (*a*); 0.99 (*b*); 0.97 (*c*).

doses of IGF-I administered intravenously as a bolus precipitate hypoglycaemia¹⁶.

There are three reasons for these differences between IGF-I and insulin. First, insulin is 50–100 times more potent than IGF-I on adipose tissue¹⁷; 5–20 times more potent on skeletal and heart muscle^{18,19} but 50–100 times less potent on chondrocytes and osteoblasts²⁰. Second, although IGF-I increases glycogen and RNA synthesis of chick embryo liver cells *in vitro*²¹ to about the same extent as insulin, an *in vivo* effect of IGF-I on glucose homeostasis at the level of the liver has not been demonstrated¹⁶. Third, IGF-I in plasma is bound to carrier proteins which compete with its receptor activity^{22,23}. Therefore, IGF-I is hypoglycaemic only when administered in amounts large enough to overcome the binding capacity of the carrier proteins allowing the factor to circulate in the free form¹⁶.

Our results demonstrate that IGF-I is the most important mediator by which growth hormone and insulin stimulate growth processes *in vivo*. It is not known how these two factors interact to induce synthesis and secretion of IGF-I by the liver. In the diabetic rat, growth hormone does not affect IGF-I levels and has no detectable anabolic effects (Fig. 1). Therefore, it does not appear to be an anabolic hormone in the absence of insulin.

The major anabolic actions of growth hormone seem to be mediated by IGF-I.

Considering the different potencies of human IGF-I and rat IGF-I^{1,14}, the rhIGF-I administered by continuous infusion appears to have the same activity as endogenously secreted rat IGF-I. Therefore, our results support the notion that IGF-I is primarily an endocrine hormone secreted mainly by the liver that reaches effector cells via the blood²⁴.

Insulin regulates the flux of substrates and their storage in liver, adipose tissue and muscle and secures the demand for substrates, whereas IGF-I triggers the long-term anabolic reactions in cells of mesodermal origin (cartilage, bone and muscle). In the future it may be possible to use IGF-I as an anabolic therapeutic agent in acute catabolic situations such as burns and polytrauma, in which administration of insulin together

with adequate parenteral nutrition does not restore anabolism because the liver is unable to secrete IGF-I in adequate amounts. The administration of IGF-I seems to offer many therapeutic advantages without involving serious risks because, first, IGF-I is a relatively weak mitogen that tends to stabilize cultured target cells in a highly differentiated state²⁵⁻²⁷; second, there are no established connections between IGF-I and any known oncogene; and third, IGF-I is present in serum of normal subjects in concentrations of 200 ng ml⁻¹, most of it bound to carrier proteins.

This work was supported by the Swiss National Science Foundation (grant 3-051-0.84). The rhIGF I was a gift of Dr W. J. Rutter (University of California, San Francisco) and Dr J. Nüesch (Ciba-Geigy, Basel). We thank Ms I. Einschenk for technical assistance and Ms M. Salmen for secretarial help.

Received 10 March; accepted 6 June 1986.

1. Froesch, E. R. & Zapf, J. *Diabetologia* **28**, 485-493 (1985).
2. Zapf, J., Schmid, Ch. & Froesch, E. R. *Clin. Endocr. Metab.* **13/1**, 3-30 (1984).
3. Eigenmann, J. E. *et al. Acta endocr.* **85**, 818-822 (1977).
4. Phillips, L. S. & Young, H. S. *Diabetes* **25**, 516-527 (1976).
5. Phillips, L. S. & Orawski, A. T. *Diabetes* **26**, 864-869 (1977).
6. Miller, L. L., Schalch, D. S. & Draznin, B. *Endocrinology* **108**, 1265-1271 (1981).
7. Cree, T. C. & Schalch, D. S. *Endocrinology* **117**, 667-673 (1985).
8. Schalch, D. S. & Cree, T. C. *Endocrinology* **117**, 2307-2312 (1985).
9. Clemmons, D. R. *et al. J. clin. Endocr. Metab.* **53**, 1247-1250 (1981).
10. Hintz, R. L., Suskind, R. & Amatayakul, K. *J. Pediatr.* **92**, 153-156 (1978).
11. Schoenle, E., Zapf, J., Humbel, R. E. & Froesch, E. R. *Nature* **296**, 252-253 (1982).
12. Greenspan, F. S., Li, C. H., Simpson, M. E. & Evans, H. M. *Endocrinology* **45**, 455-463 (1949).
13. Schlumpf, U., Heimann, R., Zapf, J. & Froesch, E. R. *Acta endocr.* **81**, 28-42 (1976).
14. Schmid, Ch., Steiner, Th. & Froesch, E. R. in *Insulin-like Growth Factors/Somatomedins* (ed. Spencer, M.) 13-29 (de Gruyter, New York, 1983).

15. Schoenle, E., Zapf, J., Hauri, Ch., Steiner, Th. & Froesch, E. R. *Acta endocr.* **108**, 167-174 (1985).
16. Zapf, J., Hauri, C., Waldvogel, M. & Froesch, E. R. *J. clin. Invest.* **77** (in the press).
17. Schoenle, E., Zapf, J. & Froesch, E. R. *Diabetologia* **13**, 243-249 (1977).
18. Poggio, C., Le Marchand-Brustel, Y., Zapf, J., Froesch, E. R. & Freychet, P. *Endocrinology* **105**, 723-730 (1979).
19. Meuli, C. & Froesch, E. R. *Eur. J. clin. Invest.* **5**, 93-99 (1975).
20. Schmid, Ch., Steiner, Th. & Froesch, E. R. *Calcif. Tissue Int.* **35**, 578-585 (1983).
21. Widmer, U., Schmid, Ch., Zapf, J. & Froesch, E. R. *Acta endocr.* **108**, 237-244 (1985).
22. Meuli, C., Zapf, J. & Froesch, E. R. *Diabetologia* **14**, 253-259 (1978).
23. Zapf, J. *et al. J. clin. Invest.* **63**, 1077-1084 (1979).
24. Schwander, J. C., Hauri, C., Zapf, J. & Froesch, E. R. *Endocrinology* **113**, 297-305 (1983).
25. Schmid, Ch., Steiner, Th. & Froesch, E. R. *FEBS Lett.* **173**, 48-52 (1984).
26. Schmid, Ch., Steiner, Th. & Froesch, E. R. *FEBS Lett.* **161**, 117-121 (1983).
27. Kurtz, A., Jelkmann, W. & Bauer, Ch. *FEBS Lett.* **149**, 105-108 (1982).

Growth factor-like action of phosphatidic acid

W. H. Moolenaar, W. Kruijer, B. C. Tilly, I. Verlaan, A. J. Bierman & S. W. de Laat

Hubrecht Laboratory, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

Phosphatidic acid (PA), an intriguing phospholipid that is rapidly produced during receptor-stimulated breakdown of phosphoinositides, has often been proposed to function as a Ca²⁺ ionophore in activated cells¹⁻⁵. The PA-ionophore hypothesis is supported by the fact that exogenously applied PA stimulates Ca²⁺ uptake in various cells⁴⁻⁶ and can evoke Ca²⁺-mediated physiological responses²⁻⁵, but it is not known whether PA accumulation affects cytoplasmic free Ca²⁺ concentration ([Ca²⁺]_i). Here we report that PA elicits a transient rise in [Ca²⁺]_i in cultured cells, not by stimulating Ca²⁺ influx, but, surprisingly, by releasing Ca²⁺ from intracellular stores. We further show that PA evokes growth factor-like effects in that it raises cytoplasmic pH, induces expression of the *c-fos* and *c-myc* proto-oncogenes and stimulates DNA synthesis. Our results indicate that, unlike an ionophore, PA acts by triggering the hydrolysis of phosphoinositides, with consequent formation of second messengers such as inositol trisphosphate signalling Ca²⁺ release. Furthermore, our data strengthen the notion that any Ca²⁺-mobilizing stimulus acting through phospholipase C may ultimately function as a growth factor.

Human A431 carcinoma cells serve as a useful system for studying changes in [Ca²⁺]_i, phosphoinositide metabolism and proto-oncogene expression in response to physiological stimuli and Ca²⁺ ionophores⁷⁻⁹. Addition of natural PA (25 µg ml⁻¹) to A431 cells loaded with the fluorescent Ca²⁺ indicator quin-2 causes an immediate, but transient, increase in [Ca²⁺]_i, whereas the Ca²⁺ ionophore A23187 induces a sustained elevation of [Ca²⁺]_i (Fig. 1a; see also ref. 9). Note that even at saturating

Table 1 Effect of Ca²⁺-mobilizing stimuli on inositol phosphate levels in A431 cells

Addition	InsP ₁	InsP ₂	InsP ₃
	(d.p.m. per sample)		
Control	1,610 ± 95	795 ± 31	875 ± 70
PA (50 µg ml ⁻¹)	2,583 ± 118	896 ± 25	1,347 ± 97
Serum (5% v/v)	2,202 ± 91	1,202 ± 91	1,455 ± 28
A23187 (0.5 µM)	1,405 ± 53	801 ± 23	789 ± 45

Levels of inositol phosphates were measured in nearly confluent cultures pre-labelled to near-isotopic equilibrium with ³H-inositol (2 µCi ml⁻¹) for at least 24 h. After washing, the cells were treated with the indicated stimuli (without Li⁺) for 30 s. The reaction was stopped by adding ice-cold trichloroacetic acid (TCA, 20% w/v). After centrifugation the supernatants were extracted with diethyl ether to remove TCA and the neutralized extracts were processed for analysis of ³H-inositol phosphates by anion-exchange chromatography as described by Berridge *et al.*²⁶. Values shown are the means ± s.e. (n = 3). We note that, as with other cell types^{27,28}, the InsP₃ fraction from A431 cells contains both the expected 1,4,5-isomer and a 1,3,4-isomer (B.C.T. *et al.*, manuscript in preparation).

concentrations of PA (50 µg ml⁻¹), the effects of A23187 and PA on [Ca²⁺]_i are not competitive, consistent with different routes of action.

Synthetic dipalmitoyl-PA also evokes a significant [Ca²⁺]_i transient (although its magnitude is more variable than that produced by natural PA), but dimyristoyl-PA and dilauroyl-PA have no effect in concentrations up to 100 µg ml⁻¹. Thus, it seems that relatively long fatty acid chains are required for an optimal PA effect. Other phospholipids, including phosphatidylcholine, phosphatidylinositol (PI) and phosphatidylserine (PS), cannot mimic the effect of PA, although a weak Ca²⁺ signal is sometimes observed with PS (50 µg ml⁻¹; data not shown). Biologically active compounds such as dioctanoylglycerol (50 µg ml⁻¹), tetradecanoylphorbolacetate (TPA, 100 nM) and fatty acids (arachidonate, oleate, stearate; 20 µM) are also ineffective in raising [Ca²⁺]_i. Furthermore, prior addition of delipidated bovine serum albumin, which would

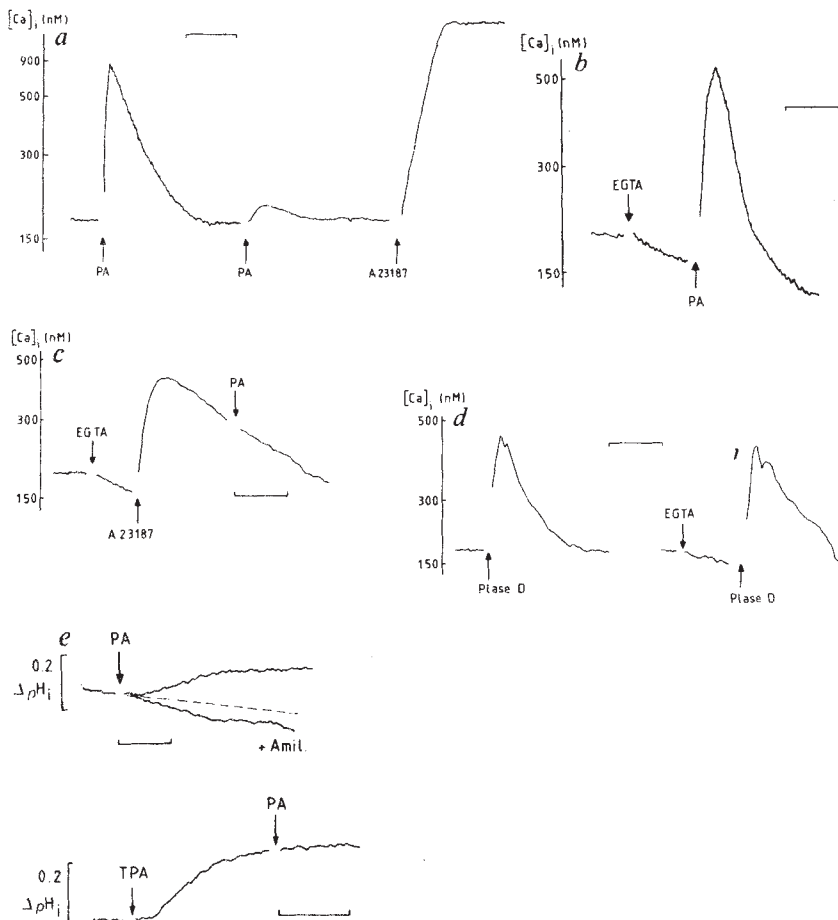


Fig. 1 *a-d*, Changes in $[Ca^{2+}]_i$ in human A431 cells after addition of natural PA and other agents as indicated. Cells were incubated for 1–2 h in serum-free medium before experimentation. Concentrations used: PA, $25 \mu\text{g ml}^{-1}$; A23187, $0.5 \mu\text{M}$; EGTA, 4 mM ; phospholipase D (from *S. chromofuscus*; Calbiochem), 0.5 U ml^{-1} . *e*, Changes in pH_i in quiescent Rat-1 cells after addition of PA ($25 \mu\text{g ml}^{-1}$) and TPA (100 nM). Amiloride (0.5 mM) was present where indicated. Cells were made quiescent by serum starvation for 24 h. Bars: *a-e*, 2 min; lower part of *e*, 5 min. **Methods:** Confluent cell cultures, attached to a rectangular glass coverslip, were loaded with either quin-2 or bis (carboxyethyl)carboxyfluorescein as described previously^{9,25}. The cells were incubated in HEPES-buffered Dulbecco's modified Eagle's medium (pH 7.40) without phenol red, and the $[Ca^{2+}]_i$ - or pH_i -dependent fluorescence was continuously monitored at 37°C using a Perkin Elmer model 3000 spectrofluorimeter. Calibration procedures and other experimental details were essentially similar to those described previously^{9,25}. Natural and synthetic L- α -PA (Na salt, 99% pure) and other phospholipids were obtained from Sigma. An aqueous dispersion of $2\text{--}5 \text{ mg ml}^{-1}$ phospholipid was briefly sonicated under N_2 and added to the incubation media to give the final concentrations indicated.

remove any free fatty acid contamination of the PA used, does not affect the Ca^{2+} response to PA (data not shown).

When extracellular Ca^{2+} is removed by EGTA, the PA-induced Ca^{2+} signal is only slightly ($<20\%$) attenuated (Fig. 1*b*), indicating that, like serum growth factors⁹, PA acts by releasing Ca^{2+} from intracellular stores rather than by stimulating the entry of Ca^{2+} across the plasma membrane. After pretreatment of the cells with A23187 in Ca^{2+} -free medium, PA cannot release any further internal Ca^{2+} stores (Fig. 1*c*), suggesting that the releasable Ca^{2+} is sequestered in membrane-enclosed organelles, presumably the endoplasmic reticulum^{9–11}, rather than bound to the inner side of the plasma membrane. Importantly, the Ca^{2+} -mobilizing action of PA is not specific for A431 cells. We observed similar $[Ca^{2+}]_i$ transients after addition of PA to Rat-1 fibroblasts and mouse embryonal mesoderm cells (data not shown).

To test whether the effects of exogenously supplied PA can be mimicked by producing PA from membrane phospholipids *in situ*, we used bacterial phospholipase D which converts phos-

pholipids to PA. A striking $[Ca^{2+}]_i$ transient is indeed observed after addition of 0.5 U ml^{-1} of phospholipase D (from *Streptomyces chromofuscus*¹²) to A431 cells, both in the presence and absence of extracellular Ca^{2+} (Fig. 1*d*). This finding strongly suggests that endogenously produced PA can serve as a potent Ca^{2+} -mobilizing stimulus.

It is generally accepted that the second messenger for releasing Ca^{2+} from internal stores is inositol-1,4,5-trisphosphate (InsP_3) produced by the phospholipase C-catalysed hydrolysis of phosphatidylinositol-4,5-bisphosphate ($\text{PtdIns}(4,5)\text{P}_2$)¹³. Table 1 shows that PA, like serum, rapidly increases InsP_3 levels in A431 cells, whereas ionophore A23187 fails to stimulate any formation of inositol phosphates. Taken together, these results indicate that PA triggers the phosphodiesteratic cleavage of inositol phospholipids and thereby mobilizes Ca^{2+} from internal stores.

As expected for a stimulus acting through phospholipase C, PA also causes the protein kinase C-mediated activation of Na^+/H^+ exchange¹⁴ in quiescent fibroblasts, as shown by an amiloride-sensitive rise in cytoplasmic pH (pH_i) of $\sim 0.15 \text{ U}$ in

Fig. 2 Induction of *c-fos* and *c-myc* mRNA expression in A431 cells after addition of natural PA ($50 \mu\text{g ml}^{-1}$) or phospholipase D (Plase D; 0.5 U ml^{-1}). Nearly confluent cells were maintained in serum-free medium for 1–2 h and then treated for 1 h with the indicated agents. C, untreated cells. RNA isolation and Northern blotting were done as described previously¹⁶.

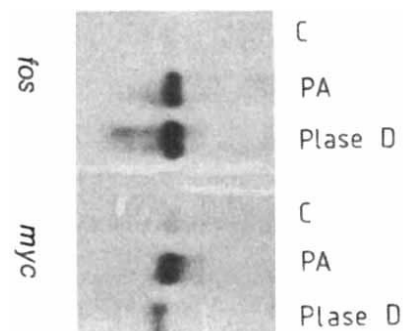


Table 2 Stimulation of DNA synthesis in quiescent Rat-1 cells

Addition	³ H-thymidine incorporation (d.p.m. × 10 ⁻¹ per well)	Stimulation (-fold)
Control	255 ± 12	1.0
Insulin (5 µg ml ⁻¹)	638 ± 25	2.5
Serum (5%, v/v)	10,715 ± 320	42.0
PA (50 µg ml ⁻¹)	3,695 ± 165	14.5
PA + insulin	8,085 ± 325	31.7
PI (50 µg ml ⁻¹)	260 ± 18	1.0
PS (50 µg ml ⁻¹)	915 ± 54	3.6
Di-octanoylglycerol (50 µg ml ⁻¹)	280 ± 15	1.1
TPA (50 nM)	560 ± 33	2.2

Confluent Rat-1 cells in 1-cm wells were made quiescent by incubation in fresh serum-free medium for 24 h. Cells were incubated for 16 h with the indicated agents before a 1-h pulse with ³H-thymidine (0.5 µCi ml⁻¹) and determination of acid-insoluble counts (means ± s.e., *n* ≥ 3).

Rat-1 cells, while addition of PA to cells prestimulated with phorbol ester (TPA) does not cause an additional elevation of *pH_i* (Fig. 1e). In contrast, ionophore A23187 does not detectably raise *pH_i* in fibroblasts (ref. 14; data not shown).

Induction of the cellular proto-oncogenes *c-fos* and *c-myc* is a striking early event in the action of various Ca²⁺-mobilizing hormones and growth factors^{8,15-17}. To determine whether PA can induce proto-oncogene expression we analysed messenger RNA from PA-treated A431 cells. The Northern blots in Fig. 2 show a significant induction of *c-fos* and *c-myc* by PA within 1 h after treatment. Moreover, exposure of the cells to phospholipase D to produce endogenous PA induces these oncogenes as strongly as exogenous PA (Fig. 2).

Finally, we investigated whether PA can mimic the effects of Ca²⁺-mobilizing hormones and growth factors on DNA synthesis in quiescent fibroblasts. Addition of PA to growth-arrested Rat-1 cells in serum-free medium causes a marked stimulation of ³H-thymidine incorporation into DNA after 16 h, whereas other phospholipids, short-chain diacylglycerol or phorbol ester (TPA) have little or no effect (Table 2). In combination with insulin, PA is nearly as potent as 5% fetal calf serum in stimulating DNA synthesis in Rat-1 cells.

Our findings are incompatible with the PA ionophore hypothesis; instead, they demonstrate that long-chain PA can trigger the hydrolysis of phosphoinositides to yield second messengers that mediate such diverse biological responses as Ca²⁺ mobilization, activation of Na⁺-H⁺ exchange, induction of proto-oncogenes and stimulation of DNA synthesis. Moreover, our data strengthen the view that any stimulus acting on the inositol phospholipid pathway may ultimately induce DNA synthesis in responsive cells.

By what mechanism does PA stimulate phosphoinositide breakdown? One possibility is that PA directly activates phospholipase C (or the putative G-protein intermediate¹⁸). A second possibility, for which there is some indirect support, is that PA causes a perturbation in the structure of the lipid bilayer, thereby rendering the PtdIns(4,5)P₂ substrate susceptible to phospholipase attack. The ability of PA to form non-bilayer domains in model membranes is well documented¹⁹⁻²², and recent studies show that phospholipase action is promoted when the bilayer is perturbed^{23,24}.

Whatever the precise mechanism, our results raise the important question of whether PA formed in the inner leaflet of the plasma membrane by diacylglycerol kinase acts in a similar manner as exogenously supplied PA to stimulate phosphoinositide hydrolysis. If so, newly synthesized PA could function in a positive-feedback loop to amplify the cascade of inositol lipid reactions following receptor stimulation. Whether this is a plausible speculation for the cellular function of PA awaits further study.

Supported by the Netherlands Cancer Foundation (Koningin Wilhelmina Fonds) and the Organisation for the Advancement of Pure Research (ZWO).

Received 16 April; accepted 24 June 1986.

1. Salmon, D. M. & Honeyman, T. W. *Nature* **284**, 344-345 (1980).
2. Putney, J. W., Weiss, S. J., Van De Walle, C. M. & Haddas, R. A. *Nature* **284**, 345-347 (1980).
3. Harris, R. A., Schmidt, J., Hitzemann, B. A. & Hitzemann, R. J. *Science* **212**, 1290-1291 (1981).
4. Ohsako, S. & Deguchi, T. *J. biol. Chem.* **256**, 10945-10948 (1981).
5. Barrit, G. J., Dalton, K. A. & Whiting, J. A. *FEBS Lett.* **125**, 137-140 (1981).
6. Ohsako, S. & Deguchi, T. *FEBS Lett.* **152**, 62-66 (1983).
7. Sawyer, S. T. & Cohen, S. *Biochemistry* **20**, 6280-6286 (1981).
8. Bravo, R., Burckhardt, J., Curran, T. & Müller, R. *EMBO J.* **4**, 1193-1198 (1985).
9. Moolenaar, W. H., Aerts, R. J., Tertoolen, L. J. G. & de Laat, S. W. *J. biol. Chem.* **261**, 279-284 (1986).
10. Prentki, M. *et al.* *Nature* **309**, 562-564 (1984).
11. Somlyo, A. P., *Nature* **309**, 516-517 (1984).
12. Imamura, S. & Horiuti, Y. *J. Biochem., Tokyo* **85**, 79-95 (1979).
13. Berridge, M. J. & Irvine, R. F. *Nature* **312**, 315-321 (1984).
14. Moolenaar, W. H. *A. Rev. Physiol.* **48**, 363-376 (1986).
15. Kelly, K., Cochran, B. H., Stiles, C. D. & Leder, P. *Cell* **35**, 603-610 (1983).
16. Kruijer, W., Cooper, J., Hunter, T. & Verma, I. M. *Nature* **312**, 711-716 (1984).
17. Müller, R., Bravo, R., Burckhardt, J. & Curran, T. *Nature* **312**, 716-720 (1984).
18. Cockcroft, S. & Gomperts, B. D. *Nature* **314**, 534-536 (1985).
19. Galla, H.-J. & Sackmann, E. *Biochim. biophys. Acta.* **401**, 509-529 (1975).
20. Papahadjopoulos, D., Vail, W. J., Pangborn, W. A. & Poste, G. *Biochim. biophys. Acta* **448**, 265-283 (1976).
21. Massari, S. & Pascolini, D. *Biochemistry* **16**, 1189-1195 (1977).
22. Verkleij, A. J., de Maagd, R., Leunissen-Bijvelt, J. & de Kruijff, B. *Biochim. biophys. Acta* **684**, 255-262 (1982).
23. Dawson, R. M. C., Irvine, R. F., Bray, J. & Quinn, P. J. *Biochem. biophys. Res. Commun.* **125**, 836-842 (1984).
24. Irvine, R. F., Letcher, A. J. & Dawson, R. M. C. *Biochem. J.* **218**, 177-185 (1984).
25. Moolenaar, W. H., Tsien, R. Y., van der Saag, P. T. & de Laat, S. W. *Nature* **304**, 645-648 (1983).
26. Berridge, M. J., Dawson, R. M. C., Downes, C. P., Heslop, J. P. & Irvine, R. F. *Biochem. J.* **212**, 473-482 (1983).
27. Irvine, R. F., Anggard, E. E., Letcher, A. J. & Downes, C. P. *Biochem. J.* **229**, 505-511 (1985).
28. Burgess, G. M., McKinney, J. S., Irvine, R. F. & Putney, J. W. *Biochem. J.* **232**, 237-243 (1985).

Normal p21^{N-ras} couples bombesin and other growth factor receptors to inositol phosphate production

Michael J. O. Wakelam*, Shireen A. Davies*,
Miles D. Houslay*, Ian McKay†, Christopher J.
Marshall† & Alan Hall†

* Molecular Pharmacology Group, Department of Biochemistry,
University of Glasgow, Glasgow G12 8QQ, UK

† Institute of Cancer Research, Chester Beatty Laboratories,
Fulham Road, London SW3 6JB, UK

Many receptors, in response to ligand activation, trigger inositol phospholipid breakdown, which leads to rapid intracellular responses¹⁻⁴. The sustained activation of this pathway is believed to be at least one of the factors involved in the stimulation of cell growth and there has been much speculation⁴⁻⁶ that certain oncogenes use this pathway to effect uncontrolled cellular proliferation. It has been suggested⁷, by analogy with the receptor-mediated control of adenylate cyclase⁸, that the receptor stimulation of inositol phospholipid metabolism is mediated through a guanine nucleotide regulatory protein (G-protein) called G_p (or N_p). Although such a species has not been identified, there is now strong experimental evidence that this process is mediated by a G-protein distinct from the stimulatory and inhibitory G-proteins (G_s and G_i, respectively)⁹⁻¹². The *ras* genes code for a plasma membrane protein, p21, whose only known biochemical property is a high-affinity GTPase activity¹³. We show here that the expression of normal p21^{N-ras} in NIH 3T3 fibroblasts leads to the coupling of certain growth factor receptors to stimulated inositol phosphate production. We propose that the *N-ras* proto-oncogene encodes a protein which couples the receptors for certain growth factors to the stimulation of phospholipase C. Thus, *N-ras* p21 may be the putative G_p or a functionally related protein.

Around 10–20% of human tumour DNA samples have been shown to contain an oncogene detectable in the NIH 3T3 DNA transfection assay^{14–16}. In almost all of these cases the oncogene is a member of the *ras* gene family^{17–19}. There are three functional *ras* genes in the human genome, *c-Ha-ras-1*, *c-Ki-ras-2* and *N-ras*, which are expressed at low levels in most normal cells. Each gene encodes a similar polypeptide, p21^{ras}, of relative molecular mass 21,000, which is found associated with the cytosolic surface of the plasma membrane²⁰; these proteins can bind GTP/GDP and express an intrinsic GTPase activity¹³. The biochemical properties and some sequence homology to the G-proteins involved in regulating adenylate cyclase and visual transduction have led to the proposal that p21 is a regulatory G-protein which mediates the effect of external growth factors²¹. To analyse the normal functioning of p21^{N-ras}, we have constructed (see Table 1 legend) an NIH 3T3 fibroblast cell line contain-

Table 1 Effect of growth factor stimulation of inositol phosphates in T15⁺ and T15⁻ NIH 3T3 cells

Ligand	Inositol phosphate (% increase over control value)	
	T15 ⁻ cells	T15 ⁺ cells
Bombesin (2.5 μ M)	12 $\pm$ 4†	134 $\pm$ 14†
GRP (1.4 μ M)	11 $\pm$ 6§	91 $\pm$ 13†
Bradykinin (3.2 μ M)	10 $\pm$ 3‡	56 $\pm$ 5†
10% calf serum	7 $\pm$ 9*	132 $\pm$ 3†
[⁸ Arg]-vasopressin (0.4 μ M)	2 $\pm$ 3*	13 $\pm$ 6§
PDGF (1.32 μ g ml ⁻¹)	4 $\pm$ 3*	12 $\pm$ 1†
EGF (0.1 μ g ml ⁻¹)	1 $\pm$ 6*	7 $\pm$ 12*
EGF + bombesin	13 $\pm$ 5	128 $\pm$ 10

T15 cells were cultured as described in Fig. 1 legend. The lipids were labelled by two different methods, neither of which altered the responsiveness of the cells. In each case the culture medium was supplemented with 2 μ Ci ml⁻¹ [³H]myo-inositol. The first labelling protocol involved a 48-h incubation in normal culture medium and the second a 16-h incubation in inositol-free DMEM containing dialysed newborn calf serum. Experiments were performed as described in Fig. 1 legend and results are expressed as means $\pm$ s.d. $n = 6$ –9 except for bombesin, where $n = 36$. Typically, the lipids in each incubation contained $\sim 80,000$ c.p.m., with $\sim 10^5$ cells, while control inositol phosphate values were $\sim 3,000$ c.p.m. Paired *t*-tests were performed and significance was assessed with regard to ligand-stimulated increases in inositol phosphates over control values for T15⁻ cells and with respect to that seen in T15⁻ cells for T15⁺ cells. As will be described elsewhere, a recombinant vector was constructed containing a normal *N-ras* gene derived from a human fetal genomic library placed downstream of the steroid-inducible MMTV-LTR promoter. Transfection into NIH 3T3 cells was achieved using the calcium phosphate co-precipitation technique. Individual colonies of cells were ring cloned and tested for their response to dexamethasone. Clone T15 showed 100% transformed colonies when treated with dexamethasone and 0% in the absence of dexamethasone.

* Not significant; † $P < 0.001$; ‡ $P < 0.01$; § $P < 0.1$; || not significant when compared with bombesin added alone.

ing the *N-ras* proto-oncogene under the control of an inducible promoter (murine mammary tumour virus long terminal repeat; MMTV-LTR). This cell line, called T15, permits us not only to switch the expression of the *N-ras* gene on and off at will but also to control the cellular concentration of normal p21^{N-ras} protein. T15 cells grown in the absence of the inducer dexamethasone (T15⁻) contain negligible human p21^{N-ras} whereas cells grown in the presence of dexamethasone (T15⁺) contain large amounts of the protein (Fig. 1). At high levels of normal p21^{N-ras} protein, corresponding to a level of inducer of 40 nM, the cells are transformed.

3T3 fibroblast cell lines exhibit receptors for a variety of growth factors, including those for bombesin, platelet-derived growth factor (PDGF) and vasopressin. Indeed, it has been shown recently in Swiss 3T3 cells that PDGF can stimulate inositol phospholipid breakdown and the generation of the two

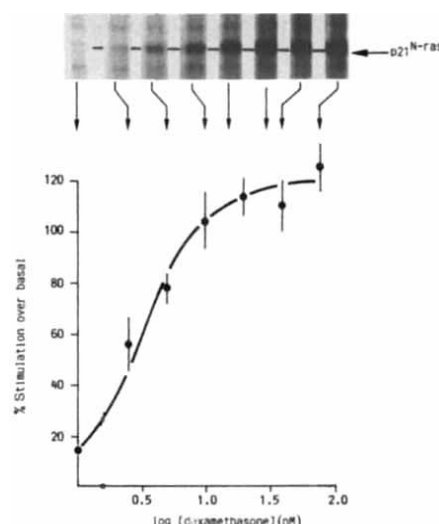


Fig. 1 Dose-response curves for the effect of dexamethasone on p21^{N-ras} expression and on bombesin-stimulated inositol phosphate accumulation in T15 cells. This figure shows an autoradiograph (top), from a typical experiment, of immunoprecipitated p21^{N-ras} from [³⁵S]-methionine metabolically labelled cells grown in the presence of the indicated concentrations of dexamethasone.

Methods. T15 cells were cultured in DMEM containing 10% (v/v) newborn calf serum and 48 h before experiments were to be performed, the medium was changed to one containing the stated concentration of dexamethasone. The cell lipids were labelled by adding [³H]myo-inositol to a final concentration of 2 μ Ci ml⁻¹ at the same time as the dexamethasone addition. After 48 h the medium was aspirated and the cells were scraped into Hank's buffered saline containing 10 mM glucose using a Teflon-coated spatula. They were washed twice in the same buffer before being dispensed into plastic insert vials containing LiCl (final concentration 10 mM) and bombesin (final concentration 2.5 μ M). Following gassing, the vials were capped and shaken in a water bath at 37 °C and 45 min later the incubations were terminated by the addition of chloroform:methanol (1:2). The phases were split by addition of chloroform and water. The inositol phosphates present in the upper phase were determined by batch chromatography on Dowex 1-X8 formate form resins^{29,30}. The [³H]-inositol lipids were determined by counting a dried sample from the lower phases. Results are expressed as means $\pm$ s.e.m. of the percentage increase in inositol phosphate production caused by bombesin, relative to control (three separate experiments $n = 9$ in each case). The rate of production of inositol phosphates was linear over the 45-min period under both basal and ligand-stimulated conditions for T15⁺, T15⁻ and wild-type cells. Immunoprecipitation³¹ of p21^{N-ras} from the cell lysate was performed using monoclonal antibody 259 which crossreacts with endogenous mouse p21^{ras} proteins as well as with human p21^{N-ras}.

second-messenger molecules, inositol 1,4,5-trisphosphate and 1,2-diacylglycerol (DAG)^{22–24}, which are responsible for elevating intracellular free [Ca^{2+}] and activating protein kinase C, respectively^{3,4}. We have used T15 cells to study the cooperation between a wide variety of growth factors and normal p21^{N-ras} in stimulating inositol phospholipid breakdown.

NIH 3T3 clone T15 cells were grown in Dulbecco's modified Eagle's medium (DMEM) containing 10% (v/v) newborn calf serum either in the absence (T15⁻) or presence (T15⁺) of dexamethasone (2 μ M unless stated otherwise). Cellular lipids were labelled with [³H]myo-inositol (Table 1 legend) for 16 or 48 h depending on culture conditions. Following removal of the cells from the dishes, and therefore from both the culture medium and radioactive label, the cells were stimulated with appropriate polypeptide and peptide growth factors. Saturating

concentrations of growth factors were used in order to obviate any changes in affinity that might have occurred between T15⁺ and T15⁻. Incubations were performed for 45 min at 37 °C in the presence of 10 mM LiCl. This methodology permitted us to determine the accumulation of inositol phosphates, which has been shown to be the most reliable and easily applied measure of receptor-stimulated inositol lipid hydrolysis^{3,4}. In transfected cells that had been cultured without dexamethasone (T15⁻), small increases in the accumulation of inositol phosphates were generated in response to growth factors (Table 1). However, in T15⁺ cells, expressing high levels of the normal *N-ras* proto-oncogene, a marked increase in inositol phosphate accumulation occurred in response to stimulation by some growth factors (Table 1). In particular, bombesin, gastrin releasing peptide (GRP), which is believed to be the mammalian form of the bombesin peptide¹⁸, and bradykinin produced a dramatic increase in inositol phosphate production in T15⁺ cells, with serum also having a stimulatory effect. The increase in inositol phospholipid breakdown was apparent only during receptor activation and was not accompanied by any increase in basal rates of turnover, which were very similar for both cells (T15⁺ = $0.042 \pm 0.010\%$ min⁻¹; T15⁻ = $0.033 \pm 0.007\%$ min⁻¹; NIH 3T3 = $0.031 \pm 0.006\%$ min⁻¹; $n = 4-9$; rate = (c.p.m. inositol phosphate produced per min $\times 100$)/c.p.m. in lipids; see Table 1 and Fig. 1 legends for experimental details). Some *ras*-expressing cells have been shown²⁵ to secrete α -transforming growth factor which binds to and activates epidermal growth factor (EGF) receptors. Table 1 shows, however, that EGF neither stimulates inositol phospholipid metabolism nor potentiates the action of bombesin in T15⁻ and T15⁺ cells.

It was recently reported that basal levels of inositol phosphates are increased in mutant *ras*-transformed NIH 3T3 cells²⁶, based on the observation of increased labelling of inositol phosphates in these cells relative to wild-type NIH 3T3 cells. It is therefore possible that mutated p21^{N-ras} might exert a direct stimulatory effect on phospholipase C independent of growth factor receptor occupancy. Indeed, our recent data using cells containing an inducible mutant *N-ras* gene (M.J.O.W. *et al.*, unpublished) support such a conclusion.

Addition of increasing concentrations of dexamethasone to T15 cells previously grown without this inducer (T15⁻) resulted in a concentration-dependent increase in expression of the human *N-ras* proto-oncogene as demonstrated by immunoprecipitable p21^{N-ras} (Fig. 1). Figure 1 shows that this progressive elevation of p21^{N-ras} concentration was paralleled by a comparable increase in the magnitude by which bombesin stimulated inositol phosphate production (Fig. 1), implying that increased efficiency of coupling of the bombesin receptor to inositol phosphate production is related to the concentration of p21^{N-ras}.

Bombesin dose-response curves for stimulated inositol phosphate production in both T15⁺ and T15⁻ cells (Fig. 2) revealed a dramatic change in the magnitude of this response for the two cell populations. However, the apparent K_a values (concentration eliciting half-maximal activation) for the stimulatory effect of bombesin were similar for both T15⁺ and T15⁻ cells (3.6 ± 0.5 and 2.8 ± 0.9 nM, respectively; $n = 3$; errors were s.d.), indicating that receptor affinity was unchanged after *N-ras* expression. To show that the marked increase in the responsiveness to bombesin seen in the T15⁺ cells was not due to any increase in receptor number, specific binding experiments were performed using ¹²⁵I-labelled GRP as a ligand (Fig. 2 legend). The number of receptors per cell was very similar, $3,192 \pm 392$ and $3,331 \pm 454$ for T15⁻ and T15⁺ cells, respectively. Binding affinities for GRP were indistinguishable, K_d values of 1.4 ± 0.1 and 1.1 ± 0.1 nM being obtained from Scatchard analysis. This is consistent with functional studies using bombesin and GRP which indicate that there is no change in the affinity of the receptors. Thus, normal *N-ras* gene expression has no effect on either the affinity or number of bombesin/GRP receptors.

We noted that wild-type NIH 3T3 cells grown in either the presence or absence of dexamethasone (2 μ M) behaved identically to the T15⁻ cells and produced only small increases in receptor-stimulated inositol phosphate production when challenged with growth factors. In wild-type NIH 3T3 cells grown either with or without dexamethasone (2 μ M) for 2 days, bombesin elicited only a $17 \pm 8\%$ and $16 \pm 8\%$ increase in inositol phosphate production. This demonstrates that it is p21^{N-ras} production and not dexamethasone itself which accounts for the enhanced coupling response.

The increase in inositol phosphate production in response to growth factor stimulation was accompanied by corresponding increases in the incorporation of ³H-thymidine into the DNA of serum-starved cells (Table 2). Such data strongly suggest a correlative link between increased inositol phospholipid breakdown and increased proliferation in the *N-ras* over-expressing cells.

Interestingly, the growth-promoting effect of EGF was enhanced in T15⁺ cells where p21^{N-ras} was expressed (Table 2). This effect was not, however, related to any marked changes in inositol phospholipid metabolism (Table 1). It is possible that this synergistic effect between EGF and elevated p21^{N-ras} concentrations is mediated by another pathway independent of inositol phospholipid turnover. Indeed, it has been reported²⁷ that EGF stimulation seems to enhance the guanine nucleotide exchange activity of p21^{N-ras}. However, an alternative explanation is possible. Several studies have shown that there is a synergistic growth-promoting effect between those growth factors which can stimulate inositol phospholipid metabolism

Table 2 Effect of growth factors on incorporation of ³H-thymidine into growth-arrested T15⁺ and T15⁻ NIH 3T3 cells

Ligands added to DMEM	T15 ⁻ cells		T15 ⁺ cells		Enhancement ratio
	Acid-insoluble radioactivity	% Increase on ligand addition	Acid-insoluble radioactivity	% Increase on ligand addition	
No added ligand	685 $\pm$ 69	~	5,065 $\pm$ 1,396	~	7.39
Bombesin (2.5 μ M)	859 $\pm$ 262	27	13,040 $\pm$ 1,223	157	15.18
Bradykinin (3.2 μ M)	823 $\pm$ 116	20	9,698 $\pm$ 2,025	92	11.78
10% Calf serum	1,463 $\pm$ 158	114	23,756 $\pm$ 1,150	369	16.20
[⁸ Arg]-vasopressin (0.4 μ M)	884 $\pm$ 152	29	8,305 $\pm$ 1,097	63	9.39
PDGF (1.32 μ g ml ⁻¹)	1,145 $\pm$ 152	67	13,824 $\pm$ 2,262	172	12.1
EGF (0.1 μ g ml ⁻¹)	745 $\pm$ 83	9	12,788 $\pm$ 1,103	152	17.2

Cells were plated in 24-well plates at a density of 10^5 cells ml⁻¹. The medium was changed 24 h later to one containing 1% (v/v) newborn calf serum and the cells were left for 5 days to become quiescent. They were then washed in DMEM and incubated in fresh DMEM, the appropriate additions made together with a final concentration of 10μ M ³H-thymidine (0.1 μ Ci μ mol⁻¹). The medium was removed 24 h later and acid-insoluble radioactivity was determined²⁸. Results are given as counts per min per 10^6 cells $\pm$ s.d. ($n = 3$). The enhancement ratio describes the increased ³H-thymidine incorporation seen in *N-ras*-expressing (T15⁺) cells as a ratio of that in 'control' (T15⁻) cells for both ligand-stimulated and basal states.

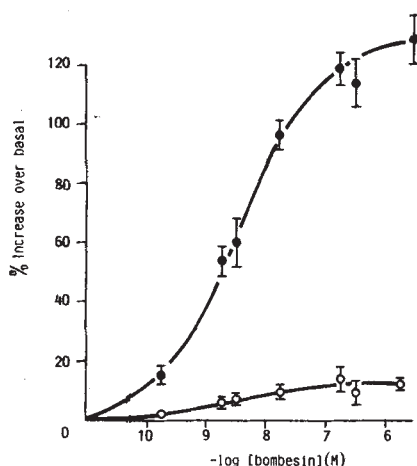


Fig. 2 Dose-response curves for the effect of bombesin on inositol phosphate accumulation in T15⁺ (●) and T15⁻ (○) cells. Cells were cultured and experiments performed as described in Fig. 1 legend. Results are means $\pm$ s.e.m., and the data are pooled from three separate experiments. Data are given as the % increase in inositol phosphate production caused by bombesin, with respect to cells incubated without bombesin.

Methods. Receptor binding studies were done using either ¹²⁵I-labelled GRP or ¹²⁵I-labelled bombesin as described elsewhere^{22,32,33}. ¹²⁵I-GRP binding was analysed by Scatchard plots and yielded results as described in the text; specific binding accounted for >80% of the total counts bound using 10⁻⁶ M unlabelled GRP to displace labelled GRP. ¹²⁵I-bombesin-specific binding yielded similar results, with an estimated 4,574 $\pm$ 191 receptors per cell for T15⁻ cells and 4,548 $\pm$ 254 receptors per cell for T15⁺ cells (errors are $\pm$ s.d. for $n=3$ separate experiments with triplicate binding assays). Specific binding was >60% of the total counts bound for ¹²⁵I-bombesin. These values are comparable to those found in pancreatic acinar cells³³ and human small cell lung cancer cells³² but are considerably less than observed for Swiss 3T3 fibroblasts²².

and EGF²⁷. We have shown (Table 1) that elevated p21^{N-ras} levels enhance inositol phosphate production following calf serum stimulation. It is therefore possible that under the conditions of the DNA synthesis assay, where 1% calf serum was present, a sufficient degree of stimulated inositol phospholipid metabolism ensued in the T15⁺ cells to allow an enhanced growth-promoting effect of EGF (Table 2). Certainly, T15⁺ cells in 1% serum have an increased rate of DNA synthesis compared with T15⁻ cells (Table 2), which presumably reflects serum-stimulated inositol phospholipid metabolism in the T15⁺ cells (Table 1).

The magnitude of the altered inositol phospholipid response to growth factors after expression of p21^{N-ras} was most evident for bombesin, GRP and bradykinin. This suggests that the action of p21^{N-ras} is not specific to bombesin/GRP receptors but rather that it amplifies the action of a spectrum of growth factor receptors capable of stimulating inositol phospholipid metabolism. We propose that normal p21^{N-ras} is a coupling protein akin, if not identical, to the putative G_p, which allows growth factor receptors to stimulate inositol lipid metabolism by activating phospholipase C. That the expression of p21^{N-ras} elicited such a marked increase in receptor-stimulated inositol phosphate generation in the particular cell type studied suggests that in normal NIH 3T3 cells the coupling protein (G_p) concentration is limiting. The expression of p21^{N-ras} appears, then, to stimulate cell growth by enhancing cellular sensitivity to stimulation by growth factors, that is, by increasing the efficiency of coupling to phospholipase C. One might well expect oncogenic mutant forms of this protein to effect transformation by increasing the basal turnover of inositol phospholipids in the absence of

exogenous growth factors rather than by enhancing receptor coupling.

This work was supported by grants from the Scottish Home and Health Department, MRC and California Metabolic Research Foundation to M.J.O.W. and M.D.H., and MRC and Cancer Research Campaign (UK) to I.M., C.J.M. and A.H.

Received 19 March; accepted 2 June 1986.

1. Michell, R. H. *Biochim. biophys. Acta* **415**, 81-147 (1975).
2. Berridge, M. J. *Biochem. J.* **220**, 345-360 (1984).
3. Berridge, M. J. & Irvine, R. F. *Nature* **312**, 315-321 (1984).
4. Downes, C. P. & Michell, R. H. *Molec. Aspects cell. Regul.* **4**, 2-56 (1985).
5. Berridge, M. J. *Biotechnology* **2**, 541-546 (1984).
6. Michell, R. H. *Nature* **308**, 770 (1984).
7. Gomperts, B. D. *Nature* **306**, 64-66 (1983).
8. Houslay, M. D. *Trends biochem. Sci.* **9**, 39-40 (1984).
9. Haslam, R. J. & Davidson, M. M. L. *FEBS Lett.* **174**, 90-95 (1984).
10. Litosch, I., Wallis, C. & Fain, J. N. *J. biol. Chem.* **260**, 5464-5471 (1985).
11. Gonzales, R. A. & Crews, F. T. *Biochem. J.* **232**, 799-804 (1985).
12. Houslay, M. D., Bojanic, D. & Wilson, A. *Biochem. J.* **234**, 737-740 (1986).
13. McGrath, J. P., Capon, D. J., Goeddel, D. V. & Levinson, A. D. *Nature* **310**, 644-649 (1984).
14. Shih, C., Paddy, L. C., Murray, M. & Weinberger, R. A. *Nature* **290**, 261-264 (1981).
15. Krontiris, T. G. & Cooper, G. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1181-1184 (1981).
16. Marshall, C. J., Hall, A. & Weiss, R. A. *Nature* **299**, 171-173 (1982).
17. Santos, E., Tronick, S. R., Aaronson, S. A., Pulciani, S. & Barbacid, M. *Nature* **298**, 343-347 (1982).
18. Pulciani, S. *et al. Nature* **300**, 539-542 (1982).
19. Hall, A., Marshall, C. J., Spurr, N. & Weiss, R. A. *Nature* **303**, 396-400 (1983).
20. Willingham, M. C., Banks-Schlegel, S. P. & Pastan, I. E. *Expl Cell Res.* **149**, 141-149 (1983).
21. Tanabe, T. *et al. Nature* **315**, 242-245 (1985).
22. Zachary, I. & Rozengurt, E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7616-7620 (1985).
23. Berridge, M. J., Heslop, J. P., Irvine, R. F. & Brown, K. D. *Biochem. J.* **222**, 195-201 (1984).
24. Hasegawa-Sasaki, H. *Biochem. J.* **232**, 99-109 (1985).
25. Heldin, C. H. & Westermark, B. *Cell* **37**, 9-20 (1984).
26. Fleischman, L. F., Chahwala, S. B. & Cantley, L. *Science* **231**, 407-410 (1986).
27. Kamata, T. & Feramisco, J. R. *Nature* **310**, 147-150 (1984).
28. Adams, R. L. P. *Expl Cell Res.* **56**, 49-54 (1969).
29. Berridge, M. J., Downes, C. P. & Hanley, M. R. *Biochem. J.* **206**, 587-595 (1982).
30. Berridge, M. J., Dawson, R. M. C., Downes, C. P., Heslop, J. P. & Irvine, R. F. *Biochem. J.* **212**, 473-482 (1983).
31. Furth, M. E., Davis, L. J., Fleurdelys, B. & Scolnick, E. M. *J. Virol.* **43**, 294-304 (1982).
32. Moody, T. W., Carney, D. N., Cutitta, F., Quattrocchi, K. & Minna, J. D. *Life Sci.* **37**, 105-113 (1985).
33. Jensen, R. T., Moody, T., Pert, C., Rivier, J. E. & Gardner, J. D. *Proc. natn. Acad. Sci. U.S.A.* **75**, 6139-6143 (1978).

Selective proteolysis defines two DNA binding domains in yeast transcription factor τ

N. Marzouki, S. Camier, A. Ruet, A. Moenne & A. Sentenac

Département de Biologie, Service de Biochimie, Centre d'Etudes Nucléaires de Saclay, 91191 Gif-sur-Yvette Cedex, France

Transcription of eukaryotic transfer RNA genes involves, as a primary event, the stable binding of a protein factor to the intragenic promoter¹⁻⁴. The internal control region is composed of two non-contiguous conserved sequence elements, the A and B blocks. These are variably spaced depending on the genes⁵⁻⁸. τ , a large transcription factor purified from yeast cells⁹, interacts with these two control elements as shown by DNase I footprinting, exonuclease digestion, dimethyl sulphate protection experiments^{9,10} and by analysis of point mutations. Here we used a limited proteolysis treatment to obtain a smaller form of τ with drastically altered DNA binding properties. A protease-resistant domain interacts solely with the B block region of tRNA genes.

Figure 1 shows the effects of various proteases on τ -DNA complexes. Complex formation between τ and the yeast tRNA_{Glu} gene was analysed by a gel retardation assay¹¹. We reasoned that any large reduction in size of τ (relative molecular mass, $M_r \sim 300,000$)⁴ by selective proteolysis should markedly alter the electrophoretic mobility of residual protein-DNA complexes. τ was first incubated with a 173-base-pair (bp) ³²P-labelled DNA fragment carrying the tRNA_{Glu} gene, and

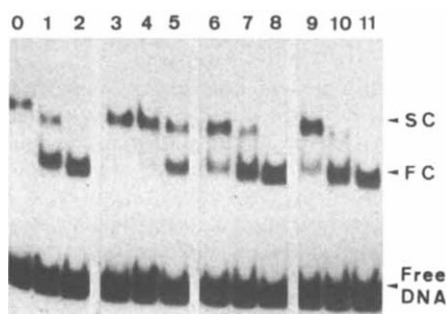


Fig. 1 Effect of various proteolytic treatments on τ -DNA complex analysed by gel electrophoresis. τ was purified by several chromatographic steps including the DNA affinity column, as previously described¹⁰. The factor was complexed with a 173-bp *Sfa*NI DNA fragment carrying the tRNA₃^{Glu} gene, in a mixture containing, in 15 μ l, 20 mM Tris-HCl pH 8.0, 1 mM EDTA, 75 mM ammonium sulphate, 80 ng carrier pBR322 plasmid DNA, 0.35 ng of the ³²P-end-labelled 173-bp DNA probe¹⁰ (3 fmol; 10,500 c.p.m.) and 20% glycerol. Binding reaction was initiated by addition of τ (1 μ l; 100 ng protein). After 15 min incubation at 25 °C, the mixtures were supplemented with different proteases as indicated and further incubated for 10 min at 25 °C. Then τ -DNA complex was separated from free DNA by gel electrophoresis on a 5% polyacrylamide gel as previously described¹². The gels were autoradiographed frozen at -70 °C with an intensifying screen. Proteases added were: lane 0, control sample, without protease; lanes 1 and 2, proteinase K (Merck) 1 and 10 ng, respectively; lanes 3-5, elastase (type II, Sigma) 100 pg, 1 and 10 ng; lanes 6-8, α -chymotrypsin (type VII, Sigma) 100 pg, 1 and 10 ng; lanes 9, 10, 11, subtilisin A (type VIII, Sigma) 100 pg, 1 and 10 ng. Slow complex (SC), fast complex (FC) and free DNA bands are indicated.

treated by different proteases at varying concentrations. The mixtures were subjected to electrophoresis and free DNA and protein-DNA complex bands were revealed by autoradiography (Fig. 1). The slow-migrating band (relative migration rate $R_f = 0.25$) seen in lane 0 corresponds to the specific binding of τ to the tRNA₃^{Glu} gene. This complex was not formed with a non-specific DNA fragment. On addition of 1 or 10 ng proteinase K a new complex was observed of $R_f = 0.5$, about twice as fast as that of the unproteolysed native complex (lanes 1, 2). Similar fast-migrating complexes were obtained with elastase (lanes 3-5), α -chymotrypsin (lanes 6-8), subtilisin A (lanes 9-11), and several other proteases. The yield of complexed DNA was not reduced by the limited protease treatments (lanes 2, 8 or 11). The abrupt transition from the slow to the fast complex suggests that the proteases had clipped off an important part of the τ molecule without greatly affecting its DNA binding affinity. Intermediate steps of proteolysis are suggested by a small increase in the migration of the slow complex (lane 1) or the fast complex (lane 2). We found that the proteolysis steps modifying the slow complex do not interfere with factor activity in a reconstituted system. Inhibition of transcription by proteolysis of τ is only correlated with the formation of the fast complex (data not shown).

We next asked whether the protein domain removed by proteolysis was involved in DNA binding or only in second factor or RNA polymerase C recognition through protein-protein contacts. A footprint analysis of protein-DNA interactions in the different complexes is shown in Fig. 2. The slow complex in the control unproteolysed sample showed the expected footprint over the entire tRNA₃^{Glu} gene (lane 2) (compare with Fig. 2 in ref. 10). The free DNA band was not protected from DNase I digestion (lane 1). This confirms the reliability and specificity of the gel DNA binding assay. Interestingly, the slow and fast complexes obtained after a limited proteolytic treatment gave markedly different footprint patterns. Whereas the slow complex

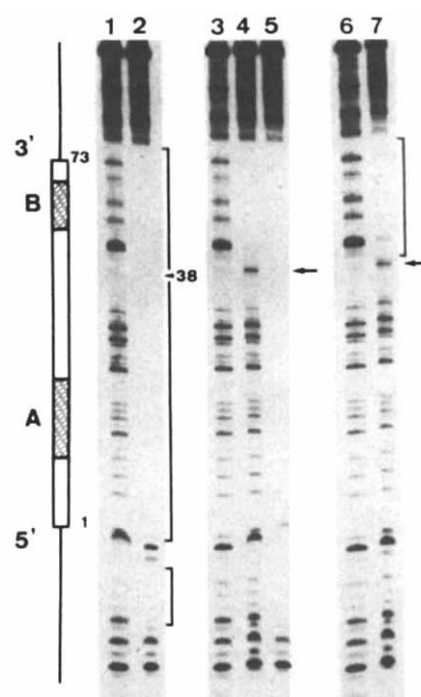


Fig. 2 Proteolysed τ shows a smaller footprint pattern on the tRNA₃^{Glu} gene in the fast-migrating complex. The binding reaction (20 μ l) between τ (100 ng protein) and the tRNA₃^{Glu} gene (15 fmol of the 173-bp DNA probe 3' end-labelled on the coding strand) was essentially as described in Fig. 1 legend, in the presence of 0.5 mM CaCl₂ and 15% glycerol for 10 min at 25 °C. Complex formation was followed by a protease treatment (with 0.1 ng or 100 ng proteinase K, for 10 min at 25 °C). Proteolysis was stopped by addition of 1 μ l 20 mM phenylmethylsulphonyl fluoride and followed by a DNase treatment in the presence of 5 mM MgCl₂ and 10 ng of DNase I for 30 s at 20 °C. The reaction was stopped by 10 mM EDTA (final concentration) and the three samples were subjected to electrophoresis on a 5% polyacrylamide gel as described in Fig. 1. The radioactive bands corresponding to the slow- and fast-migrating complexes and free DNA were located by autoradiography, excised and the DNA eluted from the gel as previously described¹². The ³²P-labelled DNA fragments were heat-denatured and analysed on a 8% sequencing gel¹². The control sample without protease is analysed in lanes 1 and 2; the sample which received 1 ng proteinase K in lanes 3-5; the sample with 100 ng proteinase K in lanes 6, 7. Lanes, 1, 3, 6, free DNA band; 2, 5, slow complex; 4, 7, fast complex. The diagram besides the autoradiograph represents the tRNA₃^{Glu} gene sequence with the A and B blocks hatched. Numbers refer to the standardized nomenclature of tRNA. The extent of DNA protection is indicated by vertical bars. Arrows point to a weak cleavage site which is unprotected in the control and enhanced in the fast complex.

retained an extended footprint similar to the control (lane 5), the fast complex had a reduced footprint, spanning only the 3' half of the gene over the B block region (lane 4). In addition, a hypersensitive site was created at position 38, in the middle of the gene. A more extensive proteolytic treatment yielding only the fast complex did not further alter its footprint pattern (lane 7). We called this protease-resistant domain τ_B as it can bind by itself to the 3' B block region of the gene. Further, these results suggest the presence of a τ_A domain interacting with the A block region. The enhanced cleavage site at position 38 defines the border of the τ_B -DNA complex. It corresponds, in the native complex, to a site that is not protected from DNase I digestion and to the most internal border accessible to λ exonuclease¹⁰. The affinity of τ_B for the tRNA₃^{Glu} gene is high. We found that the fast complex (τ_B -DNA) had the same half-life

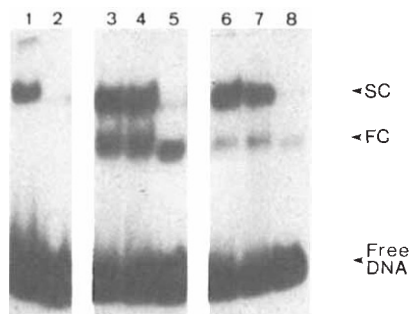


Fig. 3 Effect of single-stranded DNA on the stability of slow- and fast-migrating complexes. τ and the tRNA₃^{Glu} gene were preincubated for 10 min at 25 °C as described in the legend to Fig. 1, in the presence of 5 mM MgCl₂, 50 ng pBR322 DNA and 15% glycerol in a 20 μ l volume. The τ -DNA complex was treated with papain (0, 0.5 or 1 ng) for 10 min; then a competitor DNA was added (50 ng of native pBR322 DNA or 50 ng of single-stranded M13 DNA). After 15 min at 25 °C the complexes were analysed by gel electrophoresis as in Fig. 1 legend. The autoradiograph of the gel is shown. Papain was present in lanes 3–5 (1 ng) and 6–8 (0.5 ng); competitor pBR322 DNA was added in lanes 1, 4, 7; M13 DNA in lanes 2, 5, 8.

as the native complex when in competition with an excess of plasmid DNA containing the tRNA₃^{Glu} gene (15 min in the presence of Mg²⁺ ions and 50 min without Mg²⁺ ions).

Stillman *et al.*⁹ have shown by footprinting that single-stranded DNA interfered selectively with the interaction of τ with the 5'-region (the A block) of the tRNA₃^{Leu} gene. Using the gel binding assay, we found that addition of single-stranded M13 DNA after complex formation leads to complete dissociation of the τ -tRNA₃^{Glu} complex (Fig. 3, lane 2), whereas native pBR322 DNA had no effect. Slow and fast complexes generated by proteolysis with papain behaved differently: although the slow complex was dissociated by M13 DNA, shown by the disappearance of the band and concomitant increase in the free DNA band, the fast τ _B-DNA complex remained totally unaffected (lanes 5, 8). We concluded that the binding of single-stranded DNA to the τ _A domain destabilizes the interaction of τ _B with DNA. This destabilization, revealed by electrophoresis, suggests a strong physical interaction between τ _A and τ _B domains.

τ is a remarkable DNA-binding protein that can interact with two non-contiguous and variably spaced sequence blocks. This split promoter, as shown here, interacts with two separate protein domains of the factor molecule. These domains, which we call τ _A and τ _B (binding, respectively, to the A and B block regions), might be linked by a flexible hinge region susceptible to proteolysis, as suggested by the steep transition from the slow- to fast-migrating complexes. The protease-resistant τ _B domain can bind by itself to the 3' half of the tRNA gene but is not sufficient for gene activation. No DNA- τ _A complex has yet been isolated which may reflect the lower stability of this interaction, easily dissociated by salt¹³, single-stranded DNA⁹ or exonuclease¹⁰. The possibility of a proteolytic separation of the two domains of τ should be kept in mind while interpreting sedimentation data or complex kinetics of interaction with the split promoter. It was reported that size-fractionated τ interacted poorly with the A block region¹⁴. This could reflect the loss of the τ _A domain during sedimentation.

The dissociation of the τ -DNA complex by single-stranded DNA, as revealed by gel electrophoresis, may have some functional implications. We have proposed a model in which τ would bind alternately at the A and B regions to allow transcription of the internal promoter by RNA polymerase C¹⁰. The binding of the τ _A domain to the non-coding strand made accessible

during transcription might favour the dissociation of τ _B and transcription of the 3'-part of the gene.

The gel binding method used in this work should prove of general use for detecting the integrity of DNA-binding proteins and isolating their DNA binding core. In previous work, transcription factor TFIIA could be subdivided into functional domains by proteolytic cleavage^{15,16}. One M_r 20,000 domain recognized the 3'-side of the control region but was not transcriptionally competent by itself. By extended proteolysis this domain was shown by Miller *et al.*¹⁶ to contain repeated M_r 3,000 Zn-binding loops that are now postulated to exist in several other DNA-binding proteins^{17,18}.

Received 12 May 1986; accepted 5 June 1986.

1. Klemenz, R., Stillman, D. J. & Geiduschek, E. P. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6191–6195 (1982).
2. Lassar, A. B., Martin, P. L. & Roeder, R. G. *Science* **222**, 740–748 (1983).
3. Schaack, J., Sharp, S., Dingermann, T. & Söll, D. *J. biol. Chem.* **258**, 2447–2453 (1983).
4. Ruet, A., Camier, S., Smagowicz, W., Sentenac, A. & Fromageot, P. *EMBO J.* **3**, 343–350 (1984).
5. Galli, G., Hofstetter, H. & Birnstiel, M. L. *Nature* **294**, 626–631 (1981).
6. Ciliberto, G., Castagnoli, L., Melton, D. A. & Cortese, R. *Curr. Topics dev. Biol.* **18**, 59–88 (1983).
7. Allison, D. S., Goh, S. H. & Hall, B. D. *Cell* **34**, 655–664 (1983).
8. Dingermann, T., Sharp, S., Schaack, J. & Söll, D. *J. biol. Chem.* **258**, 10395–10402 (1983).
9. Stillman, D. J., Caspers, P. & Geiduschek, E. P. *Cell* **40**, 311–317 (1985).
10. Camier, S., Gabrielsen, O., Baker, R. & Sentenac, A. *EMBO J.* **4**, 491–500 (1985).
11. Baker, R. E., Gabrielsen, O. & Hall, B. D. *J. biol. Chem.* **261**, 5275–5282 (1986).
12. Huet, J. *et al.* *EMBO J.* **4**, 3539–3547 (1985).
13. Stillman, D. J., Sivertsen, A. L., Zentner, P. G. & Geiduschek, E. P. *J. biol. Chem.* **259**, 7955–7962 (1984).
14. Stillman, D. L., Better, M. & Geiduschek, E. P. *J. molec. Biol.* **185**, 451–455 (1985).
15. Smith, D. R., Jackson, I. J. & Brown, D. D. *Cell* **37**, 645–652 (1984).
16. Miller, J., McLachlan, A. D. & Klug, A. *EMBO J.* **4**, 1609–1614 (1985).
17. Rosenberg, U. B. *et al.* *Nature* **319**, 336–339 (1986).
18. Hartshorne, T. A., Blumberg, H. & Young, E. T. *Nature* **320**, 283–287 (1986).

A gene encoding a novel glycine-rich structural protein of petunia

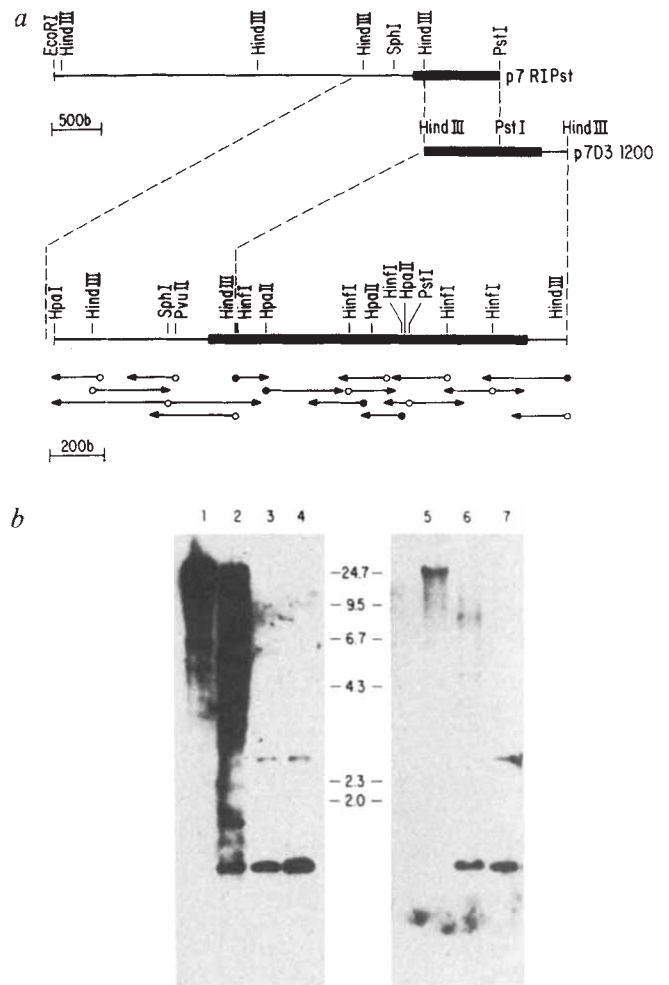
Carol M. Condit & Richard B. Meagher

Department of Genetics, University of Georgia, Athens, Georgia 30605, USA

The proteinaceous material of the cell wall of some plant species is composed of 20% or more hydroxyproline residues¹. In 1965, Lamport¹ suggested that this imino acid is derived from a hydroxyproline-rich glycoprotein (HRGP) present in cell walls. HRGP was thought to give the cell wall plasticity, thereby allowing it to grow by extension. But the cell walls of some plant species and/or organs contain very little hydroxyproline, and instead contain large amounts of glycine^{2–4}, suggesting that HRGP cannot be the only protein molecule in cell walls responsible for plasticity, and that there could be a glycine-rich cell-wall protein. Here we report the cloning and sequencing of an expressed gene from petunia, *grp-1*, the product of which is likely to function as a cell-wall structural protein in plants. The predicted amino-acid sequence of the *grp-1* gene product contains 67% glycine residues. The predicted protein structure includes a 27 amino-acid signal sequence for transport out of the cytoplasm and a 314 amino-acid region in which 90% of the residues are capable of forming a β -pleated sheet composed of 8 anti-parallel strands.

A λ clone containing the *grp-1* gene was isolated from a petunia genomic library. The coding and immediate flanking sequences of the *grp-1* gene contained on subclone p7RI-Pst and p7D3-1200 were sequenced (Fig. 1a). The sequence contains a single, continuous open-reading-frame (ORF) of 1,152 nucleotides on one strand whose nucleotide composition is 58% G but only 5% C (Fig. 2). Beginning at the initiation codon ATG (at

Fig. 1 a, Restriction endonuclease maps of subclones p7RI-Pst and p7D3-1200 and sequencing strategy. We picked up a λ genomic clone (λ 7) encoding a GRP as an anomaly during routine screening of a MG-14 partial *Sau*3A library of *Petunia hybrida* Mitchell DNA for sequences homologous with the *Ha-ras* gene⁸ and the *v-Ki-ras* gene⁹. The λ 7 hybridizes much better to the most divergent 3' sequences of both gene probes than it does to the 5' conserved regions of these genes⁸. The area of homology of λ 7 with these probes is contained on a 1,200 bp *Hind*III fragment which we subcloned into pUC18¹⁰ and named p7D3-1200. We isolated sequences 5' to and overlapping p7D3-1200 from λ 7 on a 4,000-bp *Eco*RI-Pst fragment, which we also subcloned into pUC18 and named p7RI-Pst. We mapped clones by standard procedures and performed sequencing by a modified Maxam-Gilbert¹¹ method (closed arrows) and by M13 chain-termination sequencing¹² (open arrows). **b**, Genomic Southern blot analysis of *P. hybrida* Mitchell DNA genomic sequences probed with p7D3-1200 DNA. Lanes 1, 5, 5 μ g *Bam*HI-digested genomic DNA; lanes 2, 6, 5 μ g *Hind*III-digested genomic DNA; lanes 3, 7, 1-copy reconstruction; lane 4, 2-copy reconstruction. Lanes 1-4, moderate stringency washes; lanes 5-7, high-stringency washes. Bold face numbers to the right are sizes of standards in kilobases. Genomic DNA was prepared from petunia leaves by a published procedure¹³, restriction-enzyme digested, electrophoresed on 0.8% agarose gels¹⁴, transferred to nitrocellulose and probed with ³²P-labelled, nick-translated p7D3-1200 DNA¹⁵. Hybridization was performed as previously described¹⁶. Genomic reconstruction lanes contain *Hind*III-digested p7D3-1200 and are based on a size of 1.5×10^9 bp per genomic equivalent¹⁷. The filters were washed twice for 15 min in $2 \times$ SSC, 0.2% SDS at 50 °C, and twice for 15 min in $0.1 \times$ SSC, 0.2% SDS at 50 °C (moderate stringency, lanes 1-4) or 70 °C (high stringency, lanes 5-7).



position 594 of Fig. 2) and ending at the stop codons TAATAA (at position 1,746), the ORF encodes a polypeptide composed of 384 amino-acids, 67% of which are glycine.

Analysis of the predicted amino-acid sequence of this gene reveals that the first 27 NH₂-terminal amino-acids of the predicted protein comprise an optimal signal-sequence for transport and subsequent processing⁵. The hydrophobicity plot for the protein (Fig. 4a) shows that the putative signal-sequence is followed by a hydrophilic region of 40 amino-acids which contains 13 of a total of 21 charged amino-acids present in the mature protein. The hydrophilic region is followed by a glycine-rich region (77%) comprised of two related families of repeats, each repeat containing ~40 amino-acids. The first family (F₁) has two members and the second (F₂) has five. The two families differ from each other in that the amino ends of the members of F₂ are more hydrophobic than those of F₁, and that their carboxy ends are more hydrophilic because of the presence of a histidine moiety. The carboxy terminus of the predicted protein is relatively simple, being composed of a 26 amino-acid hydrophobic region and an 11 amino-acid hydrophilic region. Thus, the presumptive mature form of this protein appears to contain three prominent structural domains: the hydrophilic amino terminus; the two families of glycine-rich repeats; and the predominantly hydrophobic carboxy terminus.

We used the plasmid p7D3-1200, which contains 90% of the coding sequence of the gene, to probe the petunia genome for related sequences. After washing genomic imprints under moderately stringent conditions, *grp-1* hybridizes to what appears to be a small multigene family, whose putative members are present at one or a few copies per genome equivalent (Fig. 1b,

lanes 1-4). When the same analysis is performed under very stringent conditions, the sequences in p7D3-1200 hybridize only to a single ~1,200-base pair (bp) *Hind*III fragment, present at one copy per genome equivalent (Fig. 1b, lanes 5-7) and comigrating with the *Hind*III insert fragment of p7D3-1200.

We next examined the steady-state messenger RNA level of the *grp-1* gene using RNA blot analysis. When we used either p7D3-1200 or single-stranded virus DNA from a M13mp9 clone containing an insert from the C-rich nonsense strand (nucleotides 1,481-1,637 of Fig. 2) as a probe, we detected four distinct homologous poly(A)⁺ RNAs of ~1,200, 1,600, 1,700 and 2,200 bases (Fig. 3a,b) in stem and leaf tissue. There is no hybridization when these RNAs are hybridized to an M13mp9 viral clone containing an insert from the G-rich coding strand (data not shown). When these Northern blots were washed under increasingly stringent conditions, there is significant hybridization with both the 1,200- and the 2,200-base RNAs, whereas hybridization to the 1,700-base RNA no longer occurs (Fig. 3c). Hybridization to the 1,600 base RNA, if present, has been obscured by shadow banding. These experiments suggest that all four RNAs potentially encode a glycine-rich protein and that the 1,700-base RNA is derived from a distant but related gene-family member, whereas the 1,200- and 2,200-base RNAs are both derived from the *grp-1* gene or closely related genes. Comparison of the amount of hybridization with the two transcriptional products of the *grp-1* gene versus that to DNA standards run in parallel reveals that these RNAs are present as ~0.02% of the total poly(A)⁺ RNA in leaf and stem tissue.

The smaller 1,200-base transcriptional product of *grp-1* could encode the large single ORF of 1,152 nucleotides found for our

1 GTAGATATTGAGGAGTTAACTCTAATATTCATGCAAGAAAGGGACACATAGAGCCT
 61 GTTTGGGTCGGCTTTTAAAGGAGCTTTTAAAGTGGCTTTTAACTCACTTATGACTTAA
 121 ACTCTAAATAAAATTAAGTTACCAACATATATTAATGGCTTAAAGCTTATAGTCAC
 181 TTTTAAACCCACCAACAGGCTCATATAAAGATGATGGCCATTTGGCATTGCCATTG
 241 CATAAAGCAATATATTTGGTTAGTGTCTGGTAGTGTGGATCAGCGTTGACTGAGTACTT
 301 TCGATCCAAAAAAGAAAGAAAGACTATAGATACAGATATATCCATTGACTTCG
 361 ACGACCAATAGGGCTATTATTTGTTCCCGCAGTTCCTCCCAACGGTCAAAACGGCTCCAA
 421 TTATACCATGATGACGCTTTGACATATGTTGGCATTATAGCTTCAGCTGACATACTA
 481 CCCTTAACCTTGTCACTTGGCTGATCGACTAACCTATATATAATTCAGAACTCTGAGCG
 541 AGGAAGGACACAAAGCAATATCTGAGAACCTTTTGTGATCATAATC ATG
 597 GGT TCT TCT CAA AAG TGG GTT ATT GGG TTA CTT CTT-TTC TCG AGC
 Gly Ser Ser Gln Lys Trp Val Ile Gly Leu Leu Leu Phe Ser Ser
 642 ATT TTC TTT GAG TTG ACT GCC ATC ACT CTT GCT GAT GAC AAG CTT
 Ile Phe Phe Gly Leu Thr Ala Ile Thr Leu Ala Asp Asp Lys Leu
 687 GAA GAG TCC AGG TGG GGT AAT GAT AAT GGG TGT GGA TTT GGT AGA
 Glu Glu Ser Arg Trp Gly Asn Asp Asn Gly Cys Gly Phe Gly Arg
 732 AGA GGT TGT GGT GGG GGT CGA TTT GGT GGA CGC GGT CCC AGT TTT
 Arg Gly Cys Gly Gly Arg Phe Gly Gly Arg Gly Pro Ser Ser
 777 GGT CGT GGT AGA GGT GCC GGA GGA TTT GGA GGT GCG GCT GGT
 Gly Arg Gly Arg Gly Ala Gly Gly Gly Phe Gly Gly Gly Ala Gly
 822 GGG GGA GCT GGC GGA GGT CTT GGT GGA GGA GGA TTA GGG GGT
 Gly Gly Ala Gly Gly Leu Gly Gly Gly Gly Gly Leu Gly Gly
 867 GGT GGA GGT GCT GGT GGA GGA GGA TTA GGG GGT GGA GGA GGT
 Gly Gly Gly Ala Gly Gly Gly Gly Gly Leu Gly Gly Gly Gly Gly
 912 GCA GGA GGA GGC TTT GGT GGT GGA GCT GGT GGA GGA GCG GGA GGT
 Ala Gly Gly Gly Phe Gly Gly Gly Ala Gly Gly Gly Ala Gly Gly
 957 GGA CTA GGT GGA GGA GGA TTA GGG GGT GGT GGA GCT GGA GGT
 Gly Leu Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
 1002 GCT GGT GGT GGA GGA GGA GTT GGT GGT GGA GCT GGT AGT GGA GGA
 Ala Gly Gly Gly Gly Gly Val Gly Gly Gly Ala Gly Ser Gly Gly
 1047 GGA TTT GGA GCT GGT GGT GGT CTT GGT GGG GGT GCT GGC CCT GGA
 Gly Phe Gly Ala Gly Gly Gly Val Gly Gly Gly Ala Gly Ala Gly
 1092 GGG GGT CTT GGA GGA GGT GGT GTC GGT GGT GGT GGA GGA GGA
 Gly Gly Val Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
 1137 GGT GTA GGT GGA GGT TCT GGT CAT GGT GGT GGA TTT GGA GCC GGA
 Gly Val Gly Gly Gly Ser Gly His Gly Gly Gly Phe Gly Ala Gly
 1182 GGA GGT GTT GGA GGT GGT GCA GGA GCG CTT GGT GCG GGA GTT
 Gly Gly Val Gly Gly Gly Gly Gly Gly Gly Leu Gly Gly Gly Val
 1227 GGT GGA GGT GGA GGA GGT AGT GGT GGT GCG GGA GGT ATA GGA
 Gly Gly Gly Gly Gly Gly Gly Ser Gly Gly Gly Gly Gly Ile Gly
 1312 GGT GGT TCA GGA CAT GGT GGA GGA TTC GGA GCC GGA GGT GGT GTA
 Gly Gly Ser Gly His Gly Gly Gly Phe Gly Ala Gly Gly Gly Val
 1317 GGT GGT GGT GTT GGA GGA GGA GCT GCA GGA GGA GGT GGT GGA GGA
 Gly Gly Gly Val Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
 1362 GGA GGC GGT GGT GGC GGA GGA GGT GGT CTT GCG GGA GGA TCA
 Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Leu Gly Gly Gly Ser
 1407 GGA CAT GGT GGT GGA TTT GGT GCA GGT GGT GGA GTT GGA GGT GGA
 Gly His Gly Gly Phe Gly Ala Gly Gly Gly Val Gly Gly Gly Gly
 1452 GCT GCT GGA GGA CTA GGT GGA GGA GGA GGA TTC GGA GGT GGT GGT
 Ala Ala Gly Gly Val Gly Gly Gly Gly Gly Gly Phe Gly Gly Gly Gly
 1497 GGT GGA GGC GTA GGG GGA GGA TCA GGT CAT GCG GGT GCG TTT GGT
 Gly Gly Gly Val Gly Gly Gly Ser Gly His Gly Gly Gly Phe Gly
 1542 GCA GGT GGA GGA GTA GGT GGT GGA GCA GGT GGA GGA TTA GGG GGA
 Ala Gly Gly Gly Val Gly Gly Gly Gly Ala Gly Gly Gly Leu Gly Gly
 1587 GGA GGA GGT GCT GGT GGA GGA GGG GGA ATA GGA GGT GGA CAT GGT
 Gly Gly Gly Ala Gly Gly Gly Gly Gly Gly Ile Gly Gly Gly His Gly
 1632 GGA GGA TTC GGA GTT GGA GTT GGA ATT GGT ATA GGA GTT GGT GTT
 Gly Gly Phe Gly Val Gly Val Gly Ile Gly Ile Gly Val Gly Val
 1677 GGT GCT GGA GCT GGA CAT GGA GTT GGC GTA GGA AGT GGA TCT GGC
 Gly Ala Gly Ala Gly His Gly Gly Gly Val Gly Ser Gly Ser Gly
 1722 AGT GGT GGC GGT GGC AAT GGC GCA TAATAAATCTGAGCCGTAGCTACT
 Ser Gly Gly Gly Gly Asn Gly Arg e
 1774 GCTCGGCTATACATATGTTGGATTTTACAATATGTTGTTAAGCTACAGCACTAA
 1834 ACAGATGAGTGTATTTCGGCTTAGCTTTTACGTACAAAATTTACTACTATATAATCTT
 1894 TATGTATCATCTTCAGATGAACTAATGAAGCTT

Fig. 2 Complete sequence of a gene encoding a GRP of petunia. Nucleotides are numbered on the left of the sequence; codon numbers are given below the predicted amino-acid sequence. a_1 , a_2 , b_1 and b_2 , the 8-bp repeated sequence GCTTAA. Repeats a_1 and b_1 contain an additional 3 bp of 5' homology, whereas repeats a_2 and b_2 contain an additional 8 bp of 3' homology. In addition, the sequences comprising repeat a_2 are contained in a 26-bp rotational palindrome, indicated by wavy lines. c and d, Other potential 5' flanking regulatory sequences. e, A potential poly(A)⁺ addition site. Underlined codons, potentially important amino-acids (see text and Fig. 4).

gene. The size of this RNA indicates that there are no introns contained in the coding sequence of this gene, and strengthens our assumption that translation begins at the ATG located at nucleotides 584–586 and ends at the TAA stop codon located at nucleotides 1,746–1,748.

We present a structure for the putative mature protein in Fig. 4b based on the assumption that the region containing seven related repeats is capable of forming an extended β -pleated sheet, as are poly-L-glycine and fibroin, a protein that consists mainly of alternating Gly-Ala residues^{6,7}. Supporting this assumption, we find that 77% of the 280 amino-acids comprising the F_1 and F_2 repeats are glycylic residues, whereas an additional 10% are the glycylic-like residues Ala and Ser, structurally interchangeable with Gly in a β -pleated sheet⁶. The non-glycylic residues in this region are non-randomly positioned and always occupy an odd-numbered position in the amino-acid sequence beginning at amino-acid 69 of GRP-1, the second amino-acid of the first F_1 member (Fig. 2). The even-numbered positions are always occupied by a Gly except at positions 250 and 288 where Gly is replaced by Ala. Thus, the entire amino-acid sequence of this 280 amino-acid region can be represented as -Gly-X-Gly-X-, where X is either Gly or one of the non-glycylic residues in this region. The similarity of the sequence -Gly-X-Gly-X- to the -Gly-Ala-Gly-Ala- of fibroin is readily apparent.

In our model (Fig. 4b) we used amino-acids 68–380 of the *grp-1* gene product to create a β -pleated sheet with 8 anti-parallel strands. We positioned the turns in the peptide backbone so that they occur at the ends of each of the seven F_1 and F_2 repeats. Placing the turns at the ends of the repeats has several advantages. It places the 5 charged histidine and 8 of the 10 phenylalanine residues present in this region on the outer edges of the sheet (Fig. 4b). Arranged in this position, the charged histidine residues would be unable to interfere with the hydrogen bonding between the anti-parallel strands. In addition, it allows us to create the eighth anti-parallel strand using the amino-acids of the carboxy-terminal hydrophobic region. This region also has the sequence Gly-X-Gly-X, where X is predominantly large amino-acids with bulky side chains. Having eight anti-parallel strands would allow this region of the molecule to assume a barrel shape, raising further intriguing functional possibilities for this protein. In our model, we allow intrachain hydrogen bonding to occur only between even-numbered amino-acids (Gly to Gly) and between odd-numbered amino-acids (X to X) but never between odd- and even-numbered amino-acids (Gly to X). This creates a structure where the R-groups of all of the non-glycine amino-acids would project in the same direction out of the plane of the sheet, thereby generating a large surface with strongly hydrophobic regions available for interaction with other hydrophobic structures such as the cell membrane. The model we present here is constructed to maximize hydrogen bonding and, therefore, the stability of the molecule. The dimensions of this molecule as a sheet would be $\sim 13 \times 3 \times 0.5$ nm. It is only one of many possible β -pleated sheet configurations this region could assume, all of which would be exceedingly stable but still flexible.

Recently, G. Cassab and J. Varner (personal communication) have isolated a glycine-rich cell-wall protein fraction from

Fig. 3 Northern RNA blot analysis of *grp-1* RNAs. *a*, Lane 1, 25 μ g poly(A)⁺ RNA; 10 μ g (*a*, lane 2; *b*, lane 1) or 2.5 μ g (*c*, lanes 1-4) poly(A)⁺ RNA were hybridized to ³²P labelled p7D3-1200 (*a*, *c*) or single-stranded viral DNA from a M13mp9⁺ clone containing nucleotides 1,481-1,637 of the C-rich strand. RNA was isolated from leaves and stems of petunia by a published procedure¹³, electrophoresed on a MOPS-formaldehyde agarose denaturing gel¹⁸ and transferred to nitrocellulose. p7D3-1200 was labelled by nick translation¹⁵, the M13 single-stranded DNA probe by chain extension using a hybridization primer (New England Biolabs) and labelled according to manufacturers instructions. Hybridization was performed as previously described¹⁶. The nitrocellulose filters were washed twice in 2 $\times$ SSC, 0.2% SDS at 52.5 °C and twice in 0.1 $\times$ SSC, 0.2% SDS at 52.5 °C (*a*, *b*), at 60 °C (*c*, lane 1), at 70 °C (*c*, lane 2), at 80 °C (*c*, lane 3) or at 90 °C (*c*, lane 4). *a*, Lane 3, 0.02 ng each of *Eco*RI-cut pBR325, pBR322, pUC18 and 0.02 ng of *Hin*FI-cut pUC18; *a*, lane 4 and *b*, lane 2, as in *a*, lane 3 but 0.2 ng of each plasmid. Numbers to the left, sizes of standards in kilobases.

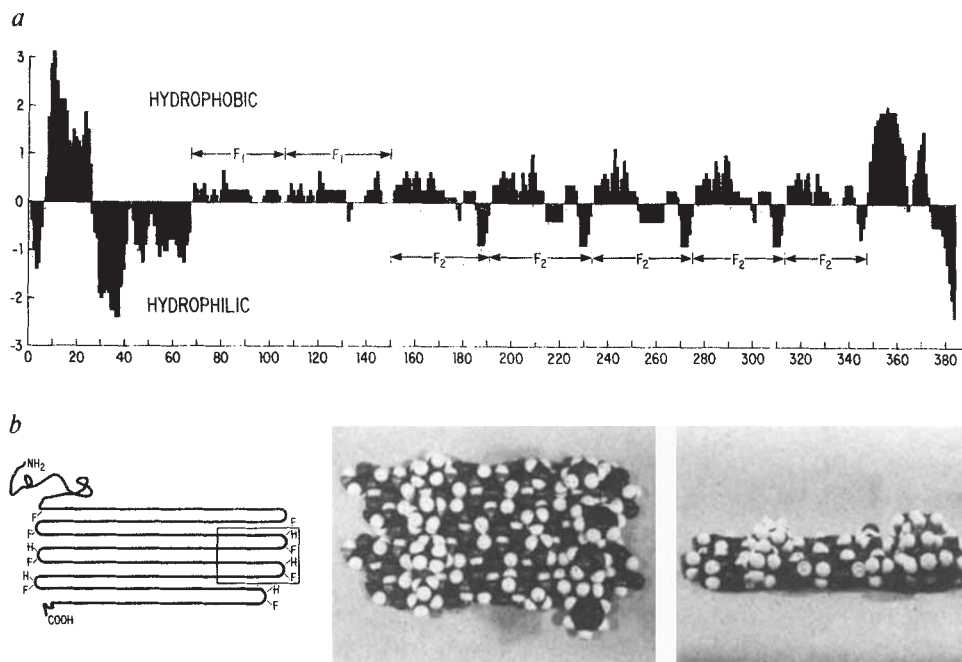
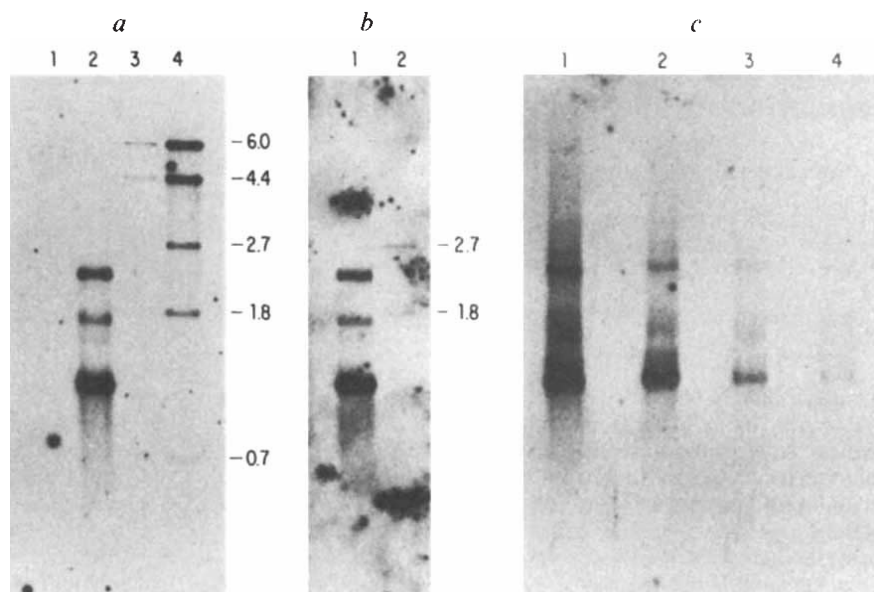


Fig. 4 *a*, Hydropathic plot of the predicted amino-acid sequence of GRP-1 (ref. 19). Ordinate, amino-acid number; abscissa, hydrophobicity. Arrows, region included in each member F₁ and F₂. *b*, Left, Model for the potential structure of the *grp-1* gene product. Box, denotes region shown in molecular model. F, phenylalanine; H, histidine. Middle and right, Space-filling model of a portion of the predicted GRP-1 β -pleated sheet. Middle, top view; right, side view. Note that in this model the histidine and phenylalanine residues are all positioned on the outer edge of the molecule and that they and all other amino-acids with large bulky side chains project upwards from the plane of the β -pleated sheet.

pumpkin seed coats with a density of 1.58 g ml⁻¹, the predicted density of the *grp-1* protein product. Their work provides further evidence that GRPs are structural cell wall components and indicates that these proteins will be found in a wide variety of plants.

We thank Dr J. Varner and G. Cassab for their encouragement and for access to unpublished data; Dr D. Shah for a λ petunia

genomic library; Dr B. Geever for help with Maxam-Gilbert sequencing and with construction of the space-filling model of *grp-1*; Dr V. Baird for petunia genomic DNA; Dr L. Pearson, L. Mendoza, Dr C. Keith and Dr G. Kishore for discussion; and T. Schurr for reading the manuscript. Supported by grants from the University of Georgia Biotechnology Project and Monsanto Corporation.

Received 28 March; accepted 23 May 1986.

1. Lampert, D. T. A. *Adv. Bot. Res.* **2**, 151-218 (1965).
2. Melin, D., Vallee, J. C., Vansuyt, G. & Prevost, J. *Physiol. Pl.* **46**, 68-72 (1979).
3. Dreher, M. L., Weber, W., Bemis, W. P. & Berry, J. W. *J. Agric. Food Chem.* **28**, 364-366 (1980).
4. Rackis, J. J., Anderson, R. L., Sasame, H. A., Smith, A. K. and Van Etten, C. H. *Agric. Food Chem.* **9**, 409-412 (1961).
5. Von Heijne, G. *Eur. J. Biochem.* **133**, 17-21 (1983).
6. Lucas, F. & Rudall, K. M. *Comp. Biochem.* **26B**, 475-558 (1968).
7. Pauling, L. & Corey, R. B. *Proc. natn. Acad. Sci. U.S.A.* **37**, 729-740 (1951).
8. Fasano, O., Taparowsky, E., Fiddes, J., Wigler, M. & Goldfarb, M. *J. molec. appl. Genet.* **2**, 173-180 (1983).

9. Ellis, R. W. *et al. Nature* **292**, 506-511 (1981).
10. Yanisch-Perro, C., Vieira, J. & Messing, J. *Gene* **33**, 103-119 (1985).
11. Shah, D. M., Hightower, R. C. & Meagher, R. B. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1022-1026 (1982).
12. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467.
13. Shure, M., Wessler, S. & Federoff, N. *Cell* **35**, 225-233 (1983).
14. Nagao, R. T., Shah, D. M., Eckenrode, V. K. & Meagher, R. B. *DNA* **1**, 1-9 (1981).
15. Rigby, P. W. J., Dreckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237-251 (1977).
16. Condit, C., Hagen, T. J., McKnight, T. D. & Meagher, R. B. *Gene* **25**, 101-108 (1983).
17. Bennett, M. D., Smith, J. B. & Heslop-Harrison, J. S. *Proc. R. Soc. B216*, 179-199 (1982).
18. Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning* 202 (Cold Spring Harbor Laboratory, New York, 1982).
19. Kyte, J. and Doolittle, R. F. *J. molec. Biol.* **157**, 105-132 (1982).

Endogenous ADP-ribosylation of G_s subunit and autonomous regulation of adenylate cyclase

C. Jacquemin*, H. Thibout*, B. Lambert† & C. Correze*

* INSERM—U.96 Unité de Recherche sur la Glande, Thyroïde et la Régulation Hormonale, 78, rue du Gl Leclerc—94270 Le Kremlin Bicêtre, France

† Department of Biochemistry, Faculté des Sciences 51062 Reims Cedex, France

Although cholera toxin induces a marked stimulation of adenylate cyclase activity in rat adipocyte plasma membranes¹, the holotoxin induces only a slight increase of cyclic AMP accumulation in intact cells². A similar apparent anomaly is seen with pertussis toxin, which has been shown to inhibit the G_i subunit of adenylate cyclase, and has a greater effect on cAMP accumulation and lipolysis than the activation by cholera toxin of the G_s subunit³. To understand better the way in which these bacterial toxins are modifying the adipocyte cells, we prepared adipocyte plasma membranes and submitted them to ADP-ribosylation by cholera and pertussis toxins. During the incubation of control cells, we found endogenous ADP-ribosylation of G_s as a result of sustained stimulation of G_i by adenosine. Our results point to a possible homeostatic system in which the autonomous adjustment of the basal activity of G_s as a function of that of G_i , under the control of feedback inhibitory ligands, ensures a steady production of cAMP within the cell.

We have previously shown³ that, when rat adipocytes are incubated for up to 3 h in the absence of adenylate cyclase stimulators and phosphodiesterase inhibitors, the level of cAMP remains constant. Here we find that the capacity of the G_s subunit of adenylate cyclase to be ADP-ribosylated by cholera toxin decreased during this period, whereas the capacity of G_i to be ADP-ribosylated by pertussis toxin (IAP) was unaltered (Fig. 1a, b). Labelled G_s components of relative molecular mass 42,000 (42K) and 46,000 (46K) were almost completely lost

when plasma membranes prepared from cells preincubated for up to 90 min, were treated with ³²P-NAD and activated cholera toxin. A similar loss of G_s ADP-ribosylation capacity was found for membranes prepared from cells preincubated with cholera toxin (Fig. 1c).

This change in G_s activity could be explained either by generation of an inhibitor or by covalent modification of G_s . If the latter were the case, endogenous phosphorylation or the occurrence of ADP-ribosylation within the cells are likely. To distinguish between these two possibilities rat adipocytes were incubated with nicotinamide, an inhibitor of ADP-ribosylation⁴ (Fig. 2a). The maximal effective concentration of nicotinamide (25 mM) not only prevented the drop in G_s labelling during incubation, but increased it by two to threefold, indicating that there is a high proportion of ADP-ribosylated species in the basal state. It also suggests that this proportion results from a dynamic equilibrium between ADP-ribosylation and the reverse reaction. This interpretation can be supported by the recent descriptions of NAD:arginine ADP-ribosyltransferase activities⁵ and of ADP-ribose-arginine cleavage enzymes⁶ in turkey erythrocytes.

The spontaneous endogenous ADP-ribosylation of G_s that occurred during preincubation of the cells could be prevented by IAP (Fig. 1d), as shown by the restored labelling of 42–46K protein when the corresponding membranes were further treated with cholera toxin. Moreover a significant labelling of the 40–41K protein appeared in these experimental conditions. During the same preincubation of the cells, G_i was ADP-ribosylated, as expected, under the action of IAP, as shown by the reduced labelling of the 40–41K species when the corresponding membranes were treated with IAP in the presence of ³²P-NAD. However, the ADP-ribosylation of G_i in IAP-treated cells does not block the ADP-ribosylation induced by cholera toxin, rather the combined ADP-ribosylation in the presence of both toxins is greater than the sum of the effects of each toxin alone. This result suggests that ADP-ribosylation of G_i by cholera toxin, recently described⁷, involves sites separate from those interacting with IAP, although both toxins decrease the action of inhibitory ligands^{2,7}. This suggests that there may be intervention of an agonist produced by the cells and acting via G_i to stimulate

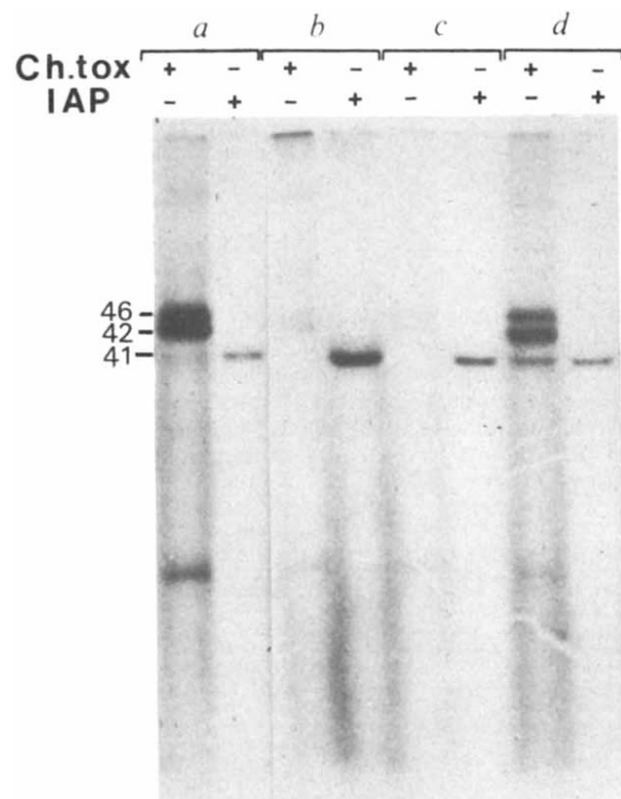


Fig. 1 SDS-polyacrylamide gel electrophoresis and autoradiographic analysis of membrane proteins labelled in the presence of cholera toxin or IAP and ³²P-NAD. Adipocytes were isolated as described in ref. 14. The washed cells (1.27 mmol triacylglycerol per assay) were incubated for 90 min in Krebs-Ringer bicarbonate buffer containing 4% bovine serum albumin. a, Control: membrane from unincubated cells; b, cells preincubated in the absence of toxin; c, cells preincubated with cholera toxin (5 μ g ml⁻¹); d, cells preincubated with IAP (10 μ g ml⁻¹).

Methods. At the end of the incubation period the washed cells were disrupted in 10 mM Tris-HCl, pH 7.4, containing 250 mM sucrose and proteases inhibitors (0.1 mM phenylmethylsulphonyl chloride, 10 μ g ml⁻¹ leupeptin and aprotinin)¹⁵. Plasma membranes were prepared as described in ref. 16. Aliquots of membrane (200 μ g protein) were ADP-ribosylated with 3 μ M ³²P-NAD (33 Ci mmol⁻¹) in the presence of activated cholera toxin (70 μ g ml⁻¹) or IAP (μ g ml⁻¹) as described in refs 17, 18. Membranes were recovered by centrifugation and the washed pellets were dissolved in 1% SDS and separated by SDS-polyacrylamide gel electrophoresis (8–13%)¹⁹. The action of the toxins on intact adipocytes was assessed by measuring the free fatty acid release (μ mol per mmol triacylglycerol) over a 90-min period. Results: control, 2.01; cholera toxin, 5.73; IAP, 15.67.

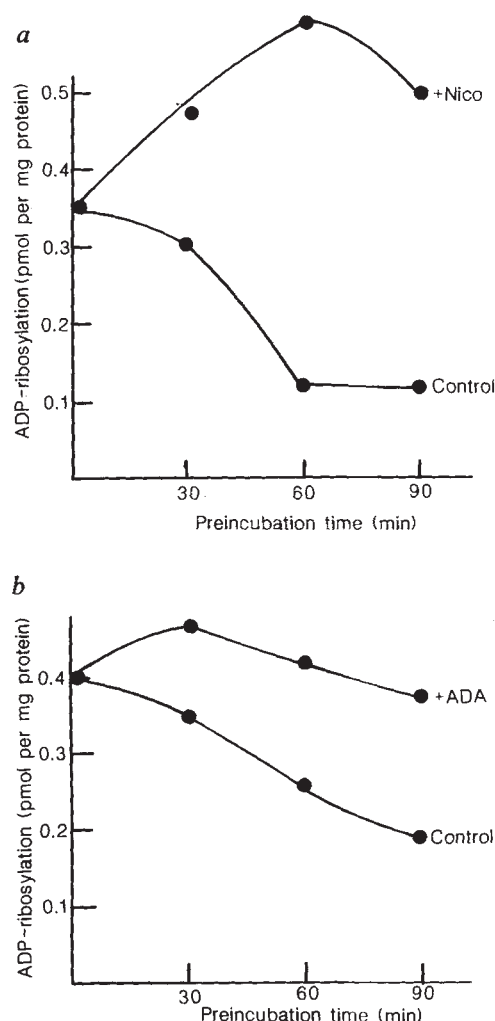


Fig. 2 Reversal of the preincubation-induced decrease in plasma membrane ADP-ribosylation by cholera toxin, by treating adipocytes with nicotinamide (*a*) or with adenosine deaminase (*b*). Adipocytes were incubated for varying periods in the absence (control) or in the presence of 25 mM nicotinamide (Nico) or adenosine deaminase (ADA) ($1 \mu\text{g ml}^{-1}$). Membranes were prepared and ADP-ribosylation was performed as in Fig. 1 in the presence of activated cholera toxin ($30 \mu\text{g per mg membrane protein}$). Plasma membrane pellets were resuspended in 40 mM Tris-HCl, pH 7.4 and aliquots ($100 \mu\text{g protein}$) were removed for radioactivity determination²⁰. The validity of this method was supported by the sole presence of labelled G_s on polyacrylamide gels (not shown).

the ADP-ribosylation of G_s . Adenosine is a possible candidate, since it is continuously formed by dephosphorylation of 5' AMP derived from cAMP and accumulates in the incubation medium⁸. Addition of adenosine deaminase to the incubation medium was found to suppress the endogenous ADP-ribosylation of G_s , suggesting that adenosine is indeed involved. In the presence of adenosine deaminase cholera toxin-stimulated ADP-ribosylation of G_s was sustained and was independent of the cell preincubation time (Fig. 2b). The defective ADP-ribosylation of membranes from adipocytes incubated for 90 min was not reported by others^{9,10} in cells incubated for 3 h. The discrepancy can be explained by variable rates of adenosine production, as these rates are highly dependent on the concentration of fat cells in the incubation media¹¹.

The question of the physiological significance of endogenous G_s ADP-ribosylation may be approached by asking (1) whether its effects are similar to those resulting from cholera toxin

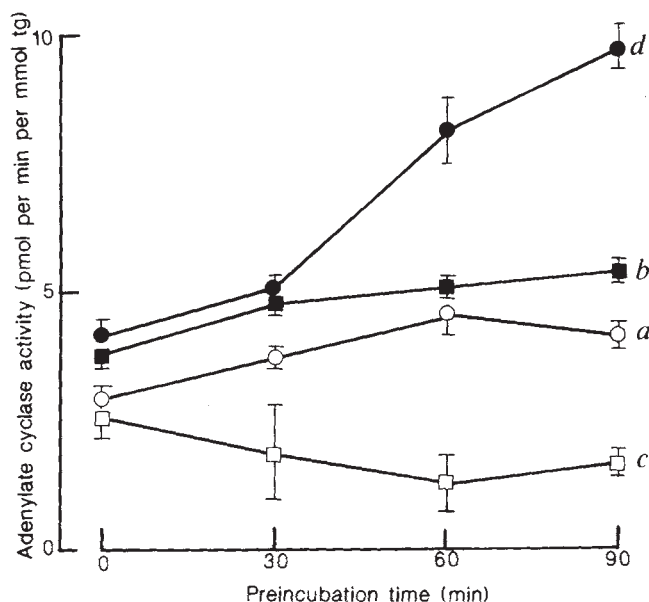


Fig. 3 Increase in adenylate cyclase activity induced by endogenous ADP-ribosylation of G_s . Fat cells were preincubated for varying periods of time as described in Fig. 1b. At the indicated times, 1 mM Ro7.2956 was added to each assay, supplemented or not with ADA ($1 \mu\text{g ml}^{-1}$) and/or Nico (25 mM), as indicated and the incubation was pursued for a further 6 min. The reaction was stopped by the addition of HClO_4 (1 M final concentration) and cAMP was assayed by a radioimmunological method¹². Adenylate cyclase activities were expressed as the mean value $\pm$ s.e.m. of triplicate experiments and reported to 1 mmol of triacylglycerol content (tg). *a*, Cells washed before 1 mM Ro7.2956 was added; *b*, cells supplemented with ADA plus Nico; *c*, cells supplemented with Nico; *d*, cells supplemented with ADA.

treatment; and (2) whether adenylate cyclase stimulation is proportional to the extent of ADP-ribosylation. The cAMP content of cells incubated for up to 3 h in control conditions remained fairly constant, indicating that its rate of synthesis by adenylate cyclase equals its rate of degradation by phosphodiesterase. When cells were incubated without changing the medium, the content of endogenous adenosine increased with incubation time⁸ and probably stimulated G_i . But despite this stimulation of G_i , the concentration of cAMP remains stable², suggesting that the positive effect of G_s increases progressively to counteract the negative effect of G_i . The experiment shown in Fig. 3 was devised to test this hypothesis.

Adipocytes were preincubated for up to 90 min in the absence of adenylate cyclase stimulators and phosphodiesterase inhibitors. At the end of the preincubation period, phosphodiesterase activity was completely suppressed by adding 1 mM Ro7.2956, which does not interfere with the adenosine receptor and allows a constant rate of cAMP accumulation for at least 20 min¹². Thus under these experimental conditions, the cAMP accumulated during a 6-min period is a measure of the adenylate cyclase activity in intact cells. The removal of endogenously produced adenosine by washing the cells showed that adenylate cyclase activity increased slightly with the preincubation time (Fig. 3a). These results were not significantly different from those obtained in experiments in which adenosine deaminase and nicotinamide were added at the end of the preincubation period to destroy adenosine and prevent ADP-ribosylation respectively (Fig. 3b). Inhibition of ADP-ribosylation by nicotinamide at the zero time of adenylate cyclase assay allowed the expression, via G_i , of the negative effect of endogenous adenosine (Fig. 3c). The destruction of adenosine by adenosine deaminase before adenylate cyclase assay suppressed the negative effect of G_i , revealing a regular increase in

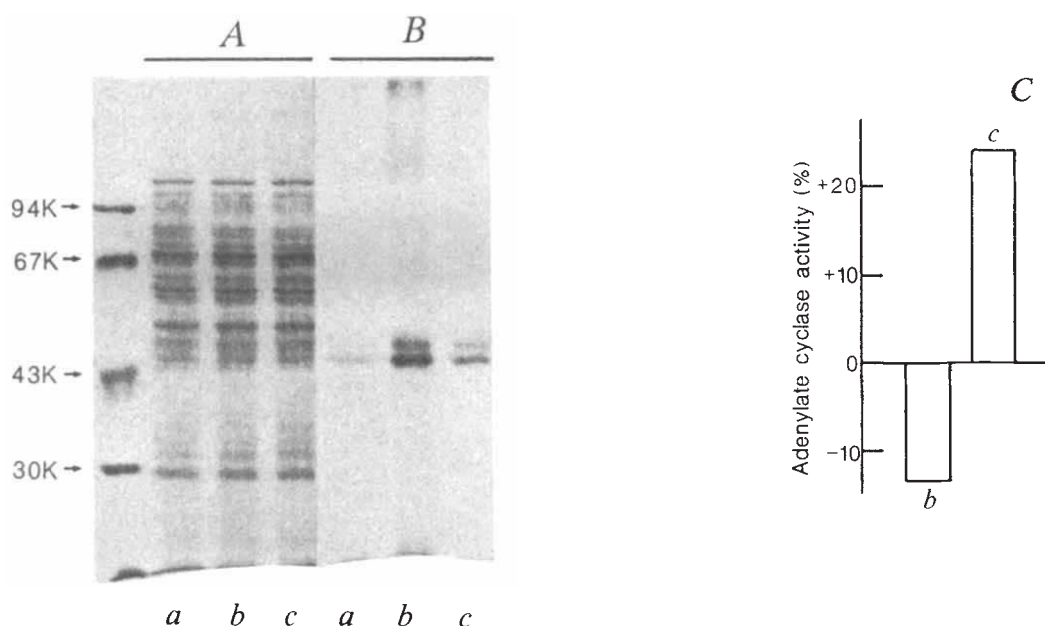


Fig. 4 Relationship between the effects of adenosine deaminase and phenylisopropyladenosine on ADP-ribosylation of G_s and adenylate cyclase activity. Cells were preincubated for 20 min in the presence of adenosine deaminase ($1 \mu\text{g ml}^{-1}$) (b) or adenosine deaminase, plus PIA ($1 \mu\text{M}$) (c) and compared with unincubated cells (a). Labelled proteins were resolved by electrophoresis as described in Fig. 1. A, Coomassie blue staining; B, autoradiographic analysis. C, Per cent variation of adenylate cyclase activity from control.

Methods. Membranes were prepared and ADP-ribosylated in the presence of ^{32}P -NAD and activated cholera toxin or assayed for adenylate cyclase activity. The final incubation mixture ($100 \mu\text{l}$) contained in addition to the membrane aliquot, 1.5 mM [$\alpha\text{-}^{32}\text{P}$]ATP ($5 \times 10^5 \text{ cpm}$), 20 mM phosphocreatine, 5 U ml^{-1} phosphocreatine-kinase, 5 mM MgCl_2 , 0.1 mM dithiothreitol, 0.1 mM GTP- $\gamma\text{-S}$. After 10 min incubation at 37°C , the reaction was stopped by adding $200 \mu\text{l}$ of a solution containing 20 mM ATP, 1 mM ^3H -cAMP ($4 \times 10^4 \text{ c.p.m.}$) and 2% SDS. The [$\alpha\text{-}^{32}\text{P}$]cAMP produced was quantified according to Salomon *et al.*²¹. Assays were carried out in triplicate. Enzyme activity is expressed as % variation from control value ($554 \pm 31 \text{ pmol cAMP per min per mg protein}$).

adenylate cyclase activity in the absence of any exogenous stimulator (Fig. 3d). These findings may be interpreted as the result of an endogenous ADP-ribosylation of G_s resulting from the chronic stimulation of G_i during the preincubation period. The observed negative effect of nicotinamide in the brief 6-min period provides further evidence of a rapid turnover rate of ADP-ribosylated and non-ADP-ribosylated forms of G_s .

To confirm the role of adenosine in the described modifications of the labelling capacity of G_s and adenylate cyclase activity, an experiment was devised in which endogenous adenosine was destroyed by adenosine deaminase and replaced by phenylisopropyladenosine (PIA). After 20 min incubation, cells were homogenized and membranes were ADP-ribosylated in the presence of cholera toxin or assayed for adenylate cyclase activity. In these conditions, the destruction of adenosine during the incubation period increased the labelling of G_s and decreased the adenylate cyclase activity (Fig. 4b). After PIA challenge, the labelling of G_s was decreased close to control and adenylate cyclase activity was stimulated by 37% in the presence of GTP- $\gamma\text{-S}$ (Fig. 4c).

Although ADP-ribosylation induced by cholera toxin and that induced by the endogenous process have similar effects on the activity of adenylate cyclase, the nature of the chemical modification involved appears to be different: the toxin effect has been reported to be irreversible¹³, whereas the endogenous process is readily reversible. This result also implies that endogenous ADP-ribosylation of G_s prevents its ADP-ribosylation by cholera toxin.

To conclude, the present results show that the sustained stimulation of G_i by adenosine enhances an ADP-ribosylating transferase acting on the G_s subunit of adenylate cyclase. This endogenous reversible modification, which prevents the irreversible modification induced by cholera toxin, increases the positive influence of G_s , resulting in a constant adenylate cyclase activity. The β - γ subunits released from G_i during its activation may be the link involved in this homeostatic regulation.

We thank Mrs A. Lefevre, Mr. M. Bahloul and Mr L. Outin for help in the preparation of the manuscript.

Received 27 January, accepted 11 July 1986.

1. Malbon, C. C. & Gill, D. M. *Biochim. biophys. Acta* **586**, 518–527 (1979).
2. Lambert, B. & Jacquemin, C. *FEBS Lett.* **143**, 188–192 (1982).
3. Lambert, B., Aublin, J. L., Champion, S., Haye, B. & Jacquemin, C. *FEBS Lett.* **181**, 390–396 (1985).
4. Katada, T. & Ui, M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3129–3133 (1982).
5. Moss, J. & Vaughan, M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3621–3624 (1978).
6. Moss, J., Jacobson, M. K. & Stanley, S. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5603–5607 (1985).
7. Owens, J. R., Frame, L. T., Ui, M. & Cooper, D. M. *J. biol. Chem.* **260**, 15946–15952 (1985).
8. Schwabe, U., Ebert, R. & Erbiler, H. C. *Adv. Cyclic Nucleotide Res.* **5**, 569–584 (1975).
9. Murayama, T. & Ui, M. *J. biol. Chem.* **258**, 3319–3326 (1983).
10. Malbon, C. C., Rapiejko, P. J. & Mangano, T. J. *J. biol. Chem.* **260**, 2558–2564 (1985).
11. Londos, C., Cooper, D. M. F., Schlegel, W. & Rodbell, M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5362–5366 (1978).
12. Lambert, B. & Jacquemin, C. *FEBS Lett.* **105**, 19–22 (1979).
13. Cassel, D. & Selinger, Z. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3307–3310 (1977).
14. Rodbell, M. *J. biol. Chem.* **242**, 5744–5750 (1967).
15. Avruch, J. & Wallach, D. F. H. *Biochim. biophys. Acta* **233**, 334–347 (1971).
16. Belsham, G. J., Denton, R. M. & Tanner, M. J. *Biochem. J.* **192**, 457–467 (1980).
17. Berthillier, G., d'Alayer, J. & Monneron, A. *Biochem. biophys. Res. Commun.* **109**, 297–304 (1982).
18. Berthillier, G., Megret, F., Alouf, J. E. & Monneron, A. *C.r. hebd. Séanc. Acad. Sci. Paris* **297**, 575–578 (1983).
19. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
20. Pfeuffer, T. *FEBS Lett.* **101**, 85–89 (1979).
21. Salomon, Y., Londos, C. & Rodbell, M. *Analyt. Biochem.* **58**, 541–548 (1974).

Erratum

The crystal structure of d(GGATGGGAG): an essential part of the binding site for transcription factor IIIA

M. McCall, T. Brown, W. N. Hunter & O. Kennard

Nature **322**, 661–664 (1986).

THE title of this letter should read as above. The version originally printed is ambiguous.

The statistics of quasar pairs

BURBIDGE, Narlikar and Hewitt¹ (BNH) have recently resurrected the notion that there may be a significant excess of close pairs of quasars (QSOs) with different redshifts. If true, this would constitute an argument against the cosmological interpretation of redshifts. BNH report such an effect, but several other studies, made over the last several years, find no such excess²⁻⁶. This discrepancy is explored here and it is concluded that the BNH sample of QSO pairs is biased by selection effects.

A decade ago, the early QSO pairs were the subject of some fascination. Discovered at a time when the total number of known QSOs was still small, they suggested to some that QSO redshifts may be non-cosmological^{7,8}, although others regarded such a conclusion to be premature⁹⁻¹¹. In 1976, Bolton *et al.*¹² reported a sample of radio-loud QSOs which contained an exceptional number of close QSO pairs. This stimulated subsequent searches made specifically to find more such QSO pairs, but these turned out to be disappointing—Wills and Wills carefully repeated the Bolton *et al.* procedure using an independent sample of the same size, but this time the result was consistent with a random distribution of QSOs¹³. This might be considered to have settled the matter, but by the early 1980s there were enough QSOs known to allow clustering analyses of large, uniform samples, and several such studies have now been made using a variety of techniques²⁻⁶. Again, no consistent evidence for significant clustering has been found.

Recently, however, in work reminiscent of that of Bolton *et al.*¹² BNH collected from the literature QSO pairs for which the angular separation (θ) is < 2 arc min and at least one member is a radio source, and, making assumptions about the parent population, they concluded that there is a highly significant excess of such pairs. Such an excess would have been easily seen in the several other studies made in recent years (which were not referenced by BNH), and the BNH work stands out in apparent contradiction to all these others.

To illustrate the magnitude of this discrepancy, Fig. 1 shows the cumulative frequency distribution of QSO pairs as a function of angular separation. This figure is based on work in an earlier paper on QSO pairs⁶, using QSOs found in objective prism and grism surveys, which should be unbiased with respect to clustering. If QSOs are distributed randomly, the cumulative number of pairs should increase as the square of the angular separation (solid line in Fig. 1), until edge effects set in.

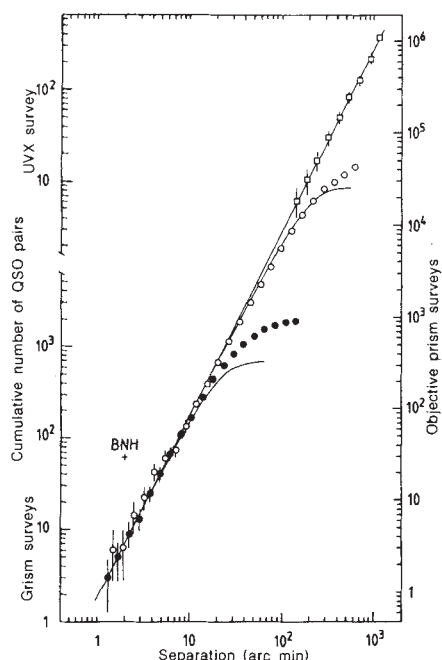


Fig. 1 Cumulative frequency distribution of QSO pairs as a function of angular separation. The open circles, filled circles, and open squares are from the objective prism, grism and UVX surveys listed, respectively, in Table 1 (a), (b) and (c) of ref. 6; the objective prism data were further supplemented by a more recent survey from ref. 5. The error bars correspond to $\sqrt{N}$. The straight line has a slope of +2, the theoretical slope for a random distribution; the curves are Monte Carlo simulations for single fields. The cross marks the position expected for the data point at $\theta = 2$ arc min if QSOs were clustered as suggested by BNH.

The points in Fig. 1 correspond closely to this expectation. At small angular separations they follow the straight line within the statistical uncertainties, and deviate from it at larger separations due to edge effects, first the grism pairs (34-arc-min fields), and then the objective prism pairs (5° fields). The turnovers at large separations are not so sharp as in the Monte Carlo simulations because some of the fields are adjacent to each other, allowing pair separations greater than the plate dimensions. The ultraviolet excess sample, covering a larger area of sky, shows that the distribution of QSOs continues to be random on scales of up to tens of degrees. Any non-random effects (reduced sensitivity at plate edges, or the actual physical clustering of QSOs which may be expected from the cosmological hypothesis and has probably now been detected¹⁴) will tend to flatten these distributions, so the results in Fig. 1 give an upper limit to any clustering of QSOs with different redshifts.

The BNH result implies an excess of close (< 2 arc min) pairs of almost an order of magnitude compared to random expectation. The cross in Fig. 1 shows

where the point for $\theta = 2$ arc min would be if such an excess applied to pairs of small separation relative to those of larger separation. The disagreement is obvious.

What is the cause of this discrepancy? It cannot be due simply to the use of radio-loud rather than radio-quiet QSOs, because, as mentioned above, other searches made around radio-loud QSOs, stimulated years ago, by Bolton *et al.*¹², were negative (ref. 13). It may perhaps have to do with the determination of the parent population of the BNH sample of pairs, or equivalently, with bias in the selection of those pairs. If many of these pairs received special attention and were published early because they are close pairs, this would introduce a strong bias of unknown magnitude in favour of QSO pairs of small angular separation. Wills¹³ has discussed such selection effects at length.

There are several pairs in the BNH list which clearly support this interpretation. Of the six pairs used by BNH to draw their conclusions, four come from the work of Bolton *et al.*¹² That work originated with the discovery of an unexpectedly large number of close QSO pairs in one particular sample; such a discovery can be used to suggest a hypothesis and stimulate *a priori* tests on other samples of QSOs (as has been done, with negative results¹³), but the same pairs should not be used *a posteriori* to test the hypothesis that they themselves suggested. Of the other two pairs used by BNH, one (1038 + 528A,B) was selected for special study and publication because it was found, during a study of the radio structure of QSO candidates, to comprise two flat-spectrum sources identified with blue stellar objects¹⁵, and the other (2217 + 087A,B) because it was "the most remarkable of 20 fields centred on radio sources with steep spectra in a search for distant clusters of galaxies"¹⁶. It is impossible to determine the magnitude of the bias these selection effects give to the results of BNH, but their very existence casts doubt on any conclusions drawn from that work.

I am grateful to Leon Lucy for discussions and a critical reading of the manuscript.

P. A. SHAVER

European Southern Observatory
Karl-Schwarzschild-Str. 2
D-8046 Garching bei München, FRG

- Burbidge, G. R., Narlikar, J. V. & Hewitt, A. *Nature* **317**, 413-415 (1985).
- Osmer, P. S. *Astrophys. J.* **247**, 762-773 (1981).
- Webster, A. *Mon. Not. R. astr. Soc.* **199**, 683-705 (1982).
- Chu, Y. & Zhu, X. *Astrophys. J.* **267**, 4-11 (1983).
- Savage, A., Trew, A. S., Chen, J.-S. & Weston, A. *Mon. Not. R. astr. Soc.* **207**, 393-404 (1984).
- Shaver, P. A. *Astr. Astrophys.* **143**, 451-454 (1985).
- Stockton, A. N. *Nature phys. Sci.* **238**, 37 (1972).
- Wampler, E. J., Baldwin, J. A., Burke, W. L., Robinson, L. B. & Hazard, C. *Nature* **246**, 203-205 (1973).

9. Bahcall, J. N. & Woltjer, L. *Nature* **274**, 22-23 (1974).
10. Burbidge, E. M., Burbidge, G. R. & O'Dell, S. L. *Nature* **248**, 568-569 (1974).
11. Gott, J. R. III & Gunn, J. E. *Astrophys. J.* **190**, L105-L108 (1974).
12. Bolton, J. G., Peterson, B. A., Wills, B. J. & Wills, D. *Astrophys. J.* **210**, L1-L3 (1976).
13. Wills, D. *Phys. Scr.* **17**, 333-337 (1978).
14. Shaver, P. A. *Astr. Astrophys.* **136**, L9-L10 (1984).
15. Owen, F. N., Wills, B. J. & Wills, D. *Astrophys. J.* **235**, L57-L60 (1980).
16. Harris, D. E., Dewdney, P. E., Costain, C. H., Butcher, H. & Willis, A. G. *Astrophys. J.* **270**, 39-47 (1983).

BURBIDGE ET AL. REPLY—In answer to Shaver's comments we wish to make the following points:

(1) We did not reference most of the work cited by him (his refs 2-5) because it refers to pairs of much wider angular separation than ours. We wish to emphasize again that it is the pairs with very small separations that should be looked for as indicators of physical associations.

(2) While we agree that Shaver's work is not consistent with ours, the results given in Burbidge *et al.*¹ are to be compared with the prediction of the numbers expected by chance made in 1974 by Burbidge *et al.*².

(3) This comment by Shaver and earlier work by him³ and by Wills⁴ describes results consistent with the non-clustering of QSOs. However, the proponents of the cosmological hypothesis must have cause to worry about results like ours which appear to be inconsistent with that hypothesis. Shaver's earlier work and the present paper do not explain the discrepancy.

G. R. BURBIDGE
J. V. NARLIKAR
A. HEWITT

Center for Astrophysics and
Space Sciences,
University of California, San Diego,
La Jolla, California 92093, USA
Tata Institute of Fundamental Research,
Bombay, India

1. Burbidge, G. R., Narlikar, J. V. & Hewitt, A. *Nature* **317**, 413-415 (1985).
2. Burbidge, E. M., Burbidge, G. R. & O'Dell, S. L. *Nature* **248**, 568-569 (1974).
3. Shaver, P. A. *Astr. Astrophys.* **143**, 451-454 (1985).
4. Wills, D. *Phys. Scr.* **17**, 333-337 (1978).

Homology between IgE-binding factor gene and endogenous retroviruses

RECENTLY, Toh *et al.*¹, reported the detection of nucleotide sequence homology between segments of a rodent IgE-binding factor (IgE-BF) gene² and a Syrian hamster intracisternal A particle (IAP) genome³. Toh *et al.* proposed that the IgE-BF gene is a hybrid gene which evolved very recently by integrating segments of genes of retroviral origin with a cellular gene coding for an IgE-BF domain¹. Our own recent studies⁴ have shown that the IgE-BF cDNA is, in fact, derived from a member of the retrovirus-like mouse IAP sequence family, with its

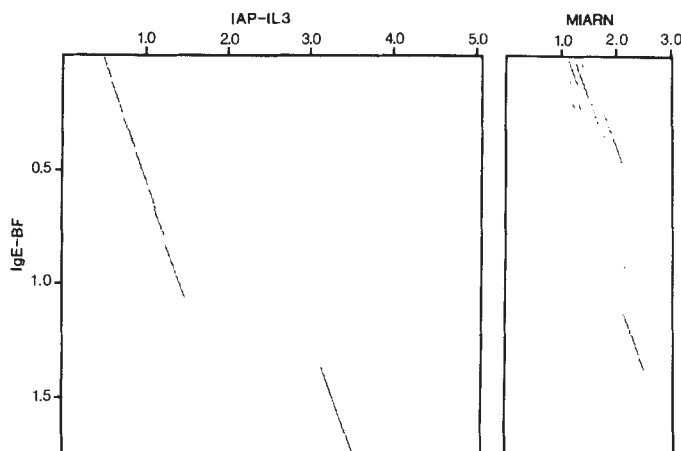


Fig. 1 Diagonal plots showing the nucleotide sequence homology between the IgE-BF coding sequence² and the murine IAP genomes IAP-IL3 (ref. 4) and MIARN (ref. 6). The comparisons were made using the program DIAGON¹⁰ with span length = 21, % = 16. The dots indicate segments of homology of 76% or greater. The scale for each of the sequences is in kilobases.

coding sequence totally derived from mouse IAP *gag* and *pol* genes.

We recently sequenced a mouse IAP genome (IAP-IL3) found in the rearranged interleukin-3 gene of the mouse myelomonocytic leukaemia cell line WEHI-3B (ref. 5). Comparison of the coding region of the IgE-BF gene with the 5.1-kb IAP-IL3 sequence⁴ showed a very high level of homology extending across the majority of the IgE-BF coding region in two major segments (Fig. 1). Most of the region of the IgE-BF coding sequence which is missing from IAP-IL3 is found in another mouse IAP element, MIARN (ref. 6) (Fig. 1). Thus, overall, 97% of the IgE-BF coding sequence is present in the two endogenous mouse IAP elements at an average homology of 90%. The restriction map of the noncoding region of the IgE-BF cDNA is very similar to that of the terminal region of IAP-IL3 suggesting that the homology extends to this region also.

The IgE-BF cDNA clone was isolated from the rat-mouse T-cell hybridoma 23B6 prepared by fusing rat T lymphocytes with cells of the AKR mouse thymoma BW5147 (ref. 7). Although one would have expected to isolate cDNA clones for rat IgE-BF from this cell line, the very high level of sequence homology with mouse IAP elements indicates that the IgE-BF cDNA was derived from a mouse IAP genome of the fusion partner. Hybridization analysis using mouse IAP gene probes has shown that mouse and rat IAP elements do not share the very high levels of homology observed between the IgE-BF gene and the mouse IAP elements⁸. The size of the cDNA clone (~3.3 kb), together with the restriction map of its 3' noncoding region suggests it was derived from a genomic RNA of a deleted IAP provirus which is intermediate in size between IAP-IL3 and MIARN. This conclusion is also in agreement with the recent findings of Moore *et al.*⁹.

The significance of this relationship between the IgE-BF gene and mouse IAP

elements is difficult to assess without further work, including the determination of whether a similar relationship also exists between IgE-BF genes and IAP-like elements in other mammals. It does, however, raise the interesting possibility that some members of the highly reiterated mouse IAP sequence family may have evolved to encode proteins with biological functions unrelated to retroviral replication.

SANIE YMER
IAN G. YOUNG

Medical Molecular Biology Unit,
John Curtin School of Medical Research,
Australian National University,
Canberra, A.C.T., 2601 Australia

1. Toh, H., Ono, M. & Miyata, T. *Nature* **318**, 388-389 (1985).
2. Martens, C. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 2460-2464 (1985).
3. Ono, M., Toh, H., Miyata, T. & Awaya, T. *J. Virol.* **55**, 387-394 (1985).
4. Ymer, S., Tucker, W. Q. J., Campbell, H. D. & Young, I. G. *Nucleic Acids Res.* **14**, 5901-5918 (1986).
5. Ymer, S., Tucker, W. Q. J., Sanderson, C. J., Hapel, A. J., Campbell, H. D. & Young, I. G. *Nature* **317**, 255-258 (1985).
6. Burt, D. W., Reith, A. D. & Brammar, W. J. *Nucleic Acids Res.* **12**, 8579-8593 (1984).
7. Huff, T. F., Uede, T. & Ishizaka, K. *J. Immun.* **129**, 509-514 (1982).
8. Lueders, K. K. & Kuff, E. L. *Nucleic Acids Res.* **9**, 5917-5930 (1981).
9. Moore, K. W. *et al. J. Immun.* **136**, 4283-4290 (1986).
10. Staden, R. *Nucleic Acids Res.* **10**, 2951-2961 (1982).

Matters Arising

Matters Arising is meant as a vehicle for comment and discussion about papers that appear in *Nature*. The originator of a Matters Arising contribution should initially send his manuscript to the author of the original paper and both parties should, wherever possible, agree on what is to be submitted. Neither contribution nor reply (if one is necessary) should be longer than 500 words and the briefest of replies, to the effect that a point is taken, should be considered.

Cassette mutagenesis shows its strength

from J. H. Richards

The hybrid mutants and mutant families created by cassette mutagenesis may advance our understanding of the relationship between function and structure.

STUDIES of the relationship between molecular structure and biological function have gained impetus from improvements in our ability to make changes, both specific and random, in the genetic material, and to observe the functional manifestations of these changes.

One approach in particular, called cassette mutagenesis¹⁻⁵, can create specific mutants or families of mutants that cannot be readily obtained by other methods of mutagenesis. This technique involves removing a stretch of DNA flanked on either end by a restriction site, then inserting a new cassette of DNA in its place. The appropriate restriction sites may be present in the gene originally or they can be introduced by site-directed mutagenesis^{1,2,6}. Hence there are few restrictions on the sequence of the inserted DNA: it can be long or short, contain many substitutions or have mixtures of bases at different sites. Four specific examples serve to highlight some of the current applications of cassette mutagenesis.

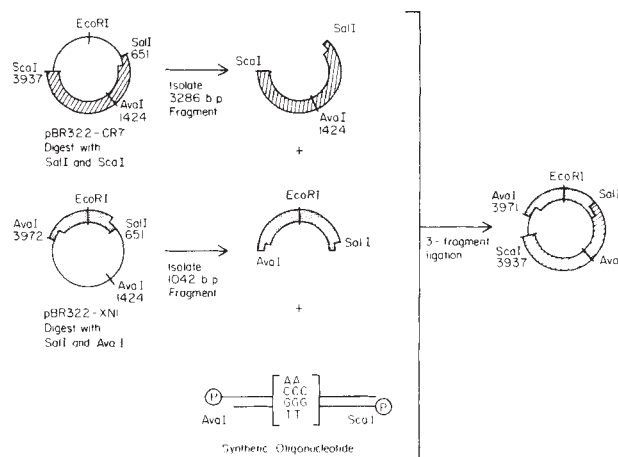
The first illustration concerns the introduction of five amino acid changes in the "outside" surface of the DNA recognition α -helix of 434 repressor with the corresponding amino acids from the recognition helix of P22 repressor⁷. The 434R and the P22R proteins bind only to their cognate operators, which have the following consensus sequences^{8,9} (• indicates a centre of symmetry and *, a non-conserved base):

A•T•AAG•••••CTT•A•T	P22 operator
ACAA•••••TTGT	434 operator

To substitute the amino acids of 434R with those present in P22R at a homologous position in the helix-turn-helix characteristic of many DNA binding proteins, a 39 bp duplex was synthesized and introduced into a vector containing the rest of the gene for 434 repressor. The resulting 434R/P22R hybrid showed the binding specificity of P22 repressor as measured both *in vivo* and *in vitro*.

Amino acid sequence homologies and three-dimensional structural similarities suggest an evolutionary kinship between β -lactamases and the D,D-carboxypeptidases-transpeptidases that cross-link the bacterial cell wall¹⁰⁻¹². To test this relationship, one objective has been to change a β -lactamase into an enzyme with detectable D,D-carboxypeptidase activity. We have used cassette mutagenesis to create a chimaeric protein containing a 30

Fig. 1 Design of three-fragment ligation for inserting the mixture of oligonucleotides in pBR322. Bp, base pair. The details of this procedure are typical of most examples of cassette mutagenesis.



amino acid insert from the *E. coli* enzyme PBP5 (MW 44,500):

50	*	78
-DLNSGKvLaeeqnadvRrdpaSlTKmmtsG-		

in place of a 29 amino acid region of RTEM-1 β -lactamase (MW 28,500):

50	*	78
-DLNSGKiLes-frpeeRFpmmStfKvllcG-		

where * denotes the active site serine (Y.H. Chang and J.H.R., in preparation). The mutant protein differs in 7 per cent of its amino acids from the parent β -lactamase and does not confer an antibiotic-resistant phenotype. But the chimaeric mutant has acquired appreciable D,D-carboxypeptidase activity, having about 3 per cent the activity of PBP5 toward diacetyl L-lys-D-ala-D-ala.

An experiment with the *CYC1* gene of *Saccharomyces cerevisiae*¹³ demonstrates the power of combining cassette mutagenesis with DNA synthesis to produce families of mutants for phenotypic screening. In studying the role of the 5' ends of mRNA in this gene, a 54-mer double-stranded synthetic oligonucleotide was inserted. During the synthesis of one strand of this duplex, one or two random mutations over a 7 bp region were introduced by using a relative amount of 0.71 M of wild-type base to 0.097 M of each of the other three bases. Primer extension was used to test the effects of these mutations on the site selection of mRNA 5' ends; a strong preference is apparent for 5' ends that align with an A residue preceded by a short tract of pyrimidine bases.

The usefulness of coupling cassette mutagenesis with the synthesis of oligonucleotide mixtures is also exemplified by a study of the role of Thr 71 in β -lactamase¹⁴. This residue is conserved in

all class A β -lactamases and lies immediately adjacent to the active site Ser at 70. All 19 possible mutants at this site were created by synthesizing two strands of DNA with essentially equimolar concentrations of A, T, G and C at the first and second, and of G and C at the third, positions of the codon for residue 71.

The resulting duplexes were incorporated into a modification of pBR322 (Fig. 1). From 108 of the resulting colonies, mutants with all 20 amino acids at residue 71 were characterized phenotypically. Surprisingly, cells producing any of 14 of the mutant β -lactamases displayed appreciable resistance to ampicillin; however, the mutants are less stable to cellular proteases than is wild-type enzyme. The results suggest that Thr 71 is not essential for binding or catalysis but plays an important role in the stability of the β -lactamase protein. □

John H. Richards is the professor of Organic Chemistry at the Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA.

- Smith, M. A. *Rev. Genet.* **19**, 423-462 (1985).
- Botstein, D. & Shortle, D. *Science* **229**, 1193-1201 (1985).
- Murphy, M.H. & Baralle, F.E. *Nucleic Acids Res.* **11**, 7685-7700 (1983).
- Lo, K.-M., Jones, S.S., Hackett, N.R. & Khorana, H.G. *Proc. natn. Acad. Sci. USA* **81**, 2285-2290 (1984).
- Buell, G., Schulz, M.-F., Selzer, G., Chollet, A. & Moura, N. *Nucleic Acids Res.* **13**, 1923-1938 (1985).
- Dalbadie-McFarland, G. & Richards, J.H. *A. Rep. Med. Chem.* **18**, 237-245 (1984).
- Wharton, R.P. & Ptashne, M. *Nature* **316**, 601-605 (1985).
- Wharton, R.P., Brown, E.L. & Ptashne, M. *Cell* **38**, 361-369 (1984).
- Potete, A.R., Ptashne, M., Balivet, M. & Eisen, H.J. *J. molec. Biol.* **137**, 81-91 (1980).
- Kelly, J.A. *et al. Science* **231**, 1429-1431 (1986).
- Samraoui, B. *et al. Nature* **320**, 376-380 (1986).
- Tipper, D.J. & Strominger, J.L. *Proc. natn. Acad. Sci. USA* **59**, 1133-1141 (1965).
- McNeil, J.B. & Smith, M. *Molec. cell. Biol.* **5**, 3545-3551 (1985).
- Schultz, S. & Richards, J.H. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1588-1592 (1986).

While the delegates to Nature's Boston conference explore the human genome, DNA manipulation dominates the exhibit area.

ADVERTISEMENTS

These notes are compiled by Karen Wright from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal and apply only within the country indicated.